STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000 page 5/140
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

423 Identification and Mutational Analysis of Major Epitopes of the Shrimp Allergen Pen a 1 (Tropomyosin) *R Ayuso, G Reese, SM Leong-Kee, MJ Plante, SB Lehrer* Tulane University Medical Center, New Orleans, LA, USA

The major allergen of shrimp, identified as the muscle protein tropomyosin, belongs to a family of highly conserved proteins with multiple isoforms found in both muscle and non-muscle cells of all species of vertebrates and invertebrates. Allergenic tropomyosins are found in invertebrates such as crustaceans (shrimp, lobster, crab, crawfish), arachnids (house dust mites), insects (cockroaches), and mollusks (e.g. squid) whereas vertebrate tropomyosins are considered to be non-allergenic. Five major IgE-binding regions of Pen a 1 (region 1: 43-57, region 2: 85-105, region 3: 133-153, region 4: 187-201 and region 5: 247-284) were identified using synthetic, overlapping peptides (15 amino acid long, offset of 6 residues) and sera of 18 shrimp-allergic subjects. The minimal epitopes (shortest peptides with maximal IgE antibody reactivity) of three (regions 1, 3, and 5) of the five major IgE binding regions were characterized. Each subject showed an individual reactivity pattern; not all epitopes were detected by all subjects. In region 1, one IgE-binding epitope, epitope 1 (43-55) was recognized with maximal IgE antibody reactivity by all four region 1-reactive subjects. Within region 3, two epitopes were identified centered around the cores 3a (137-141) and 3b (144-151) with different lengths (8 - 15 amino acids) recognized in different subjects (n = 7). In region 5, three epitopes were identified: 5a, 5b, and 5c.

Epitope 5a is centered around a common core (251-259), with a little variability in the length of the epitopes recognized by individual subjects. Epitope 5b (266-273) is identical for all 4 region 5-reactive subjects and epitope 5c (273-281) was recognized only by one subject. The effects of amino acid substitutions on the IgE binding capacity of identified epitopes were tested by comparing IgE antibody reactivities of shrimp-allergic subjects to identified Pen a 1 epitopes with those of homologous sequences of various non-allergenic vertebrate tropomyosins. In addition, all possible substitution combinations that resulted from the differences between Pen a 1 epitopes and the homologous vertebrate sequences of up to five substitutions per peptide were tested. One or two substitutions may eliminate, reduce, not alter or even increase the IgE antibody reactivity whereas more than two substitutions abolished the IgE binding ability of modified Pen a 1 peptides. In general, substitutions in the center of the epitope sequence are more likely to eliminate (59.5%) IgE binding as compared to substitutions in the peripheral parts of the epitopes (39.1%). Most of the single substitutions that eliminate IgE binding are non-conservative (63.5%) as compared to 23.1% of conservative substitutions. Thus, single amino acid non-conservative substitutions in a core part of an epitope are most likely to significantly reduce or abolish IgE antibody binding and may be useful in future development of safer allergen vaccines or less allergenic foods.

424 Cross-Reactivity of Walnut and Other Tree Nuts by IgE Immunoblot Inhibition *SS Teuber*†, SK Sathe‡§, KH Roux‡, WR Peterson** *University of California, Davis, School of Medicine, Davis, CA †Veterans Affairs Northern California Healthcare System, Mather, CA ‡Florida State University, Tallahassee, FL

Clinically, patients with walnut food allergy often report hypersensitivity to more than one nut. In our experience, this often includes pecan, which is in the same botanical family as walnut. In some patients with life-threatening reactions to walnuts, the clinical reaction to other nuts is much less severe (localized itching of the mouth, for instance), but others report reactions of the same severity. We wished to investigate the cross-reactivity between tree nuts by IgE immunoblot inhibition in order to see if there were generalizations that could be made about particular tree nuts and the potential to cross-react with walnut. Sera from patients with life-threatening walnut food allergy were used in Western IgE immunoblotting against walnut. The sera were diluted so that IgE binding to walnut proteins was visible, but not intense, maximizing the potential for other allergen sources to absorb out specific IgE. Sera were then incubated overnight with increasing concentrations of walnut extract (the homologous positive control), ovalbumin (negative control), almond, pecan, cashew and hazelnut extracts. Sera were then used in standard IgE immunoblotting to walnut protein extract separated by SDS-PAGE under reducing conditions (albumin/globulin fraction). The immunoblots showed no inhibition of specific IgE by the negative control, ovalbumin. Significantly however, pecan extract was able to absorb out IgE to walnut on a basis equivalent to that of walnut itself, at the lowest concentration of 1 mcg/mL. For each of the other nuts though, much variability in the capacity of other nut proteins to absorb out walnut-specific IgE was seen. In summary, pecan appears to be totally cross-reactive with walnut at even the lowest concentrations used, and this correlated with the clinical reports of cross-reactivity in these patients.

425 Peanut Protein Digestibility: A Gastric and Intestinal Model *CL Sellers*†, SS Teuber‡, BB Buchanan*, L Chen** *University of California at Berkeley, Department of Plant and Microbial Biology, Berkeley, CA †Cerro Cosos College, Ridgecrest, CA ‡University of California, Davis, School of Medicine

Peanut allergy is typically life-long and can be a cause of fatal anaphylaxis. The purpose of this study was to determine the digestibility of peanut allergens. Crude protein extracts of peanut were exposed to pepsin in simulated gastric fluid (SGF). The digests were subjected to SDS-PAGE and IgE immunoblotting using sera from patients with life-threatening reactions to peanuts. Allergenic proteins of the crude extract which were not digested by pepsin in SGF were isolated by gel filtration chromatography. Pepsin-resistant allergens were reduced with dithiothreitol (DTT) and exposed to pepsin in SGF. Pepsin-resistant isolates, in the oxidized state, were also exposed to pancreatin in simulated intestinal fluid (SIF). Aliquots were taken and digestion was stopped at 0, 0.25, 1, 2, 4, 15, 30, and 60 minutes. Reduced proteins exposed to pepsin were subjected to SDS-PAGE. Proteins exposed to pancreatin were subjected to SDS-PAGE, then transferred to nitrocellulose for IgE immunoblotting. Results: Three low molecular weight proteins, 14, 17 and 18 kDa were the only proteins in the crude extracts resistant to pepsin digestion. N-terminal sequencing revealed the 18 and 17 kDa proteins to be the peanut 2S protein, Ara h 2. The 14 kDa band was found to consist of multiple proteins. Ara h 2 was highly resistant to pepsin in its oxidized form, remaining intact in the gel after 30 min of exposure. The 14 kDa band displayed even greater resistance, remaining in the gel after 60 min of exposure. Reduction of Ara h 2 and the 14 kDa proteins rendered the proteins digestible by pepsin after 1 min of exposure. Ara h 2 and the 14 kDa proteins were digested by pancreatin in SIF as assessed by IgE immunoblotting by 2 min and 15 min, respectively. In summary, Ara h 2 and the 14 kDa fraction were the peanut allergens most resistant to pepsin digestion, but were digested in SIF, though not rapidly. Interestingly though, Ara h 1, a 63 kDa major allergen, was digested after a mere 15 seconds of exposure to pepsin in SGF.

426 Identification of Tropomyosin as Mollusk Allergens in Abalone, Scallop and Mussel *PSC Leung*, SH Wong†, KH Chu†* *Division of Rheumatology, Allergy and Clinical Immunology, School of Medicine, University of California, Davis, CA, USA †Department of Biology, The Chinese University of Hong Kong, Shatin, Hong Kong, China

Shellfish is one of the most common food allergens. Recently, tropomyosin was identified as the major heat stable crustacean allergen in shrimp, lobster and crab. In addition to crustaceans, many species of mollusks are commonly consumed all over the world. Interestingly, a significant portion of shellfish allergic populations is also hypersensitive to mollusks. Several mollusk allergens including *Tod p 1* in squid, *Hal m 1* in abalone and *Cra g 1* in oyster have been characterized using biochemical techniques. The objective of the present study is to determine whether tropomyosin is a candidate

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000. page 5184
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

554 Structure Plays a Critical Role in the Allergenicity of Food Proteins *GA Bannon*†, RA Kopper*†, SM Maleki, DS Shin*†, HA Sampson‡§, AW Burks‡‖* *Department of Biochemistry and Molecular Biology †University of Arkansas for Medical Sciences ‡Department of Pediatrics §Mt Sinai School of Medicine

In the past decade there has been an increase in allergic reactions to peanut proteins, sometimes resulting in fatal anaphylaxis. The development of treatment strategies for peanut allergies requires a better understanding of the immune response to these proteins and a better understanding of the structural aspects of the allergens. One of the classic characteristics of food allergens is that they are resistant to degradation during processing and digestion. In order to discover the reasons food allergens can resist degradation, the structure of the major peanut allergens, Ara h 1 and Ara h 2, were studied in relation to their IgE-binding epitopes. Ara h 1 is a protein belonging to the vicilin family of seed storage proteins while Ara h 2 is a conglutin-like seed storage protein. Both allergens are recognized by serum IgE from > 90% of peanut-sensitive patients. Ara h 1 was shown to form a highly stable homo-trimer that was stabilized by hydrophobic interactions. A molecular model of the Ara h 1 trimer was constructed to view the stabilizing hydrophobic residues in the three dimensional structure. Hydrophobic amino acids that contribute to trimer formation are at the distal ends of the 3-D structure where monomer-monomer contacts occur. Coincidentally, the majority of the IgE-binding epitopes are also located in this region suggesting that they may be protected from digestion by the monomer-monomer contacts. Upon incubation of Ara h 1 with digestive enzymes, various protease resistant fragments containing multiple IgE-binding sites were identified. These peptide fragments were protected from digestion for up to 3 hrs. On the other hand, Ara h 2 does not appear to form any stable higher order oligomeric structures. However, numerous intra-molecular disulfide bridges stabilize its tertiary structure. Just as observed with the Ara h 1 protein, the native form of the Ara h 2 allergen is also highly resistant to digestion with enzymes commonly found in the human gastrointestinal tract. When the Ara h 2 protein is reduced, thus breaking the disulfide bonds and destroying the tertiary structure of the protein, it becomes highly susceptible to digestion. In addition, when the Ara h 2 disulfide bonds are allowed to re-form in a random manner, the allergen is digested at the same rate as the reduced form of this protein. Collectively, these results show that the tertiary and quaternary structure of these two food allergens contribute significantly to their resistance to digestion, a classic characteristic of a protein destined to be a food allergen.

555 Identification and Characterization of IgE Binding Epitopes of Patatin, a Major Food Allergen of Potato *James Astwood*, Murtaza Alibhai*, Thomas Lee*, Roy Fuchs*, Hugh Sampson†* *Monsanto Company †Mount Sinai Hospital

Patatin has recently been identified as the 42 kDa major food allergen of potato (Sol t 1). Although potato allergy is rare, we have identified five potato allergic patients with positive skin prick tests and positive DBPC food challenges. Symptoms ranged from atopic dermatitis to food anaphylaxis in these patients, many of whom have been hospitalized numerous times after eating potato. Serum from these patients selectively identified and verified patatin as a major allergen from potato tubers by immunoblotting patatin protein purified from plants or total protein from potato tuber extracts. All patients indentified the 42 kDa patatin protein. The genes for patatins, a large gene family composed of numerous isoforms and which represent ~40% of potato tuber protein, have been isolated many years ago. The availability of DNA and protein sequences allowed us to utilize positive IgE binding from each patient to identify the IgE binding epitopes of patatin. By synthesizing 89 overlapping decamer oligopeptides representing the full patatin primary amino acid sequence, we have identified the major and minor IgE binding epitopes of this allergen. Subsequently, we have identified within these epitopes, substitutions in amino acid residues that appear to reduce or abolish IgE binding to these peptides. By identifying amino acid sequences that potentially reduce IgE binding, it may be possible to engineer patatins with reduced allergenicity for use in immunotherapy and other applications.

556 Macadamia Nut Anaphylaxis: Demonstration of Specific IgE Reactivity and Partial Cross Reactivity With Hazelnut *Michael Sutherland, Robyn O'Hehir, Daniel Czarny, Cenk Suphioglu* Department of Allergy, Asthma and Clinical Immunology, Monash University Medical School, The Alfred Hospital, Prahran, Victoria, Australia

BACKGROUND: Anaphylaxis due to nut allergy is a serious and potentially fatal condition. There have been no reported cases of anaphylaxis due to macadamia, an Australian tree nut originating from *Macadamia intergrifolia, Macadamia tetraphylla* and their hybrids. We report a case of anaphylaxis due to macadamia nut ingestion and confirmation of the causative allergen by immunoblotting.

AIM: To confirm macadamia nut as the causative allergen and characterise its protein and allergenic profile.

CLINICAL SUMMARY: An 18 year old female with no history of nut allergy experienced anaphylaxis after eating a flourless orange cake made with macadamia meal. On review one month later, skin tests to raw macadamia nut were positive (30 mm wheal).

METHODS: Extracts of raw and roasted macadamia nut, peanut and hazelnut were resolved on SDS-PAGE and proteins stained with Coomassie Brilliant Blue (CBB). Nut extracts were also electroblotted to nitrocellulose membranes and incubated with the patient and non allergic control sera before detection with rabbit anti-human IgE and peroxidase labelled mouse anti-rabbit IgG. Inhibition immunoblots were performed against macadamia extract with pre-incubation of the allergic patient sera (1:60 dilution) with 25 µg, 50 µg and 100 µg of macadamia, peanut and hazelnut extracts and 25 µg of macadamia oil extract.

RESULTS: CBB staining showed macadamia to have a different protein profile to the other nut allergens. The immunoblot experiments showed strong IgE binding of the patient (but not the control) serum to a 17.4 kDa protein in the macadamia extract (raw and roasted), with no significant binding to peanut and some binding to hazelnut. The inhibition immunoblots showed IgE binding was abolished by pre-incubation with macadamia extract, unaffected by pre-incubation with peanut and weakly inhibited by 100 µg of hazelnut and 25 µg of macadamia oil extract.

CONCLUSIONS: Macadamia nut can be a cause of anaphylaxis. A 17.4kDa protein present in both raw and roasted macadamia nut extracts is the principal allergen in this patient. Hazelnut and macadamia nut oil exhibit some IgE binding capacity and should also be avoided by macadamia allergic patients.

Financial support from the National Health and Medical Research Council, the Australian Allergy Foundation and The Alfred Research Trusts is acknowledged.

557 IgE Antibodies to Patatin of Potato Tuber (Sol t 1) in Atopic Infants Suspected for Potato Allergy *Heli Majamaa*, Ulla Seppälä†‡, Kristiina Turjanmaa*, Timo Palosuo‡, N Kalkkinen†, T Reunala** *Department of Dermatology, University Hospital of Tampere, Finland †Institute of Biotechnology, University of Helsinki, Finland ‡National Public Health Institute, Helsinki, Finland

Clinical manifestations of food allergy are most commonly observed within first three years of life. Allergy to potatoes is considered fairly uncommon although potatoes are widely consumed as a staple and one of the first solid foods encountered by infants. When

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

Linear IgE epitope mapping of the English walnut (*Juglans regia*) major food allergen, Jug r 1

Jason M. Robotham, BS,[a] Suzanne S. Teuber, MD,[b] Shridhar K. Sathe, PhD,[c] and Kenneth H. Roux, PhD[a] *Tallahassee, Fla, and Davis, Calif*

Background: Peanut and tree nut allergies can be life-threatening, and they appear to be growing in prevalence. Jug r 1, a 2S albumin seed storage protein, was previously characterized as a major English walnut food allergen.

Objective: We sought to identify the linear IgE-binding epitopes of Jug r 1 and to determine which, if any, amino acids are necessary for this binding to occur.

Methods: Pools of sera from walnut-allergic patients and overlapping peptides synthesized on an activated cellulose membrane were used to screen for IgE-binding epitopes. Mutational analysis of the immunodominant epitope was carried out through single and multisite amino acid substitutions. Inhibition assays were performed through use of affinity-purified IgE, soluble forms of the epitope peptide, and the recombinant 2S albumin, rJug r 1.

Results: One immunodominant linear epitope was identified. Amino acid mutations to the epitope demonstrated that the residues RGEE, at positions 36 through 39, were minimally required for IgE binding. Probing of this epitope with sera from each of 20 patients revealed 15 of the sera to be positive. Binding of patients' IgE to the epitope was inhibited with a soluble form of the peptide; however, soluble peptide did not completely inhibit the binding of IgE to the intact rJug r 1.

Conclusion: One major linear IgE-reactive epitope and its critical core amino acid residues have been identified. Mutation of any of these core amino acids resulted in loss of IgE binding to the epitope, and this points toward the feasibility of reducing allergenicity in genetically modified walnuts. However, strong evidence for the existence of conformational epitopes was also obtained. (J Allergy Clin Immunol 2002;109:143-9.)

Key words: Epitope mapping, 2S albumin, walnut, nut allergy, Jug r 1, food allergy

Abbreviations used
DMF: N,N-dimethylformamide
Fmoc: 9-fluorenylmethoxy carbonyl-derived
NC: Nitrocellulose
RT: Room temperature
TBS: Tris-buffered saline [solution]
TBS-T: Tris-buffered saline [solution] with 0.2% Tween-20

It is estimated that up to 8% of children less than 3 years of age and 2% of adults are affected by food allergies.[1] Although food-induced allergic reactions are the most common cause of outpatient anaphylaxis,[2] most severe reactions of this kind are caused by peanuts and tree nuts.[3,4] Most plant food allergens can be found among pathogenesis-related proteins, seed storage albumins and globulins, and α-amylase and protease inhibitors.[5] A previous study by Teuber et al[6] documented the cloning and sequencing of a gene encoding Jug r 1, a major allergen in the English walnut, *Juglans regia*. Jug r 1 is a 2S albumin seed storage protein and possesses important homologies in amino acid sequence with other 2S albumin proteins from Brazil nut, cottonseed, castor bean, and mustard.[6] Like many 2S albumins, Jug r 1 is synthesized as a precursor protein and cleaved into a large subunit and a small subunit, these being joined by disulfide bridges. It is currently believed that patients with life-threatening allergies to walnuts and other tree nuts will rarely become tolerant of these foods[7] and consequently face a lifetime of avoidance.[8] Avoidance, however, is often difficult because of the ubiquitous nature of these foods and the potential for cross-contamination during their processing.[7]

Many approaches to allergen immunotherapy under investigation today are based on a detailed knowledge of the amino acids found in IgE-reactive B-cell epitopes.[9-12] These reactive sites can be either linear or conformational. Typically, a linear epitope contains a stretch of contiguous amino acids spanning 5 to 10 residues in the antigen, whereas residues distant in the primary sequence but proximate in the folded protein constitute a conformational epitope.[13] In recent years, the genes encoding several proteins with allergenicity, such as Jug r 1, have been cloned and expressed through use of molecular biology techniques.[6,14-16] Despite the rapidly increasing number of recombinant allergens, relatively few IgE-reactive B-cell epitopes have been defined.[17-29]

From [a]the Department of Biological Science and Structural Biology Program, Florida State University, Tallahassee; [b]the Department of Internal Medicine, Division of Rheumatology, Allergy and Clinical Immunology, University of California, Davis, School of Medicine; and [c]the Department of Nutrition, Food, and Exercise Science, Florida State University, Tallahassee.

Funded in part by the International Life Sciences Institute, the Allergy and Immunology Institute (Suzanne S. Teuber), USDA NRICGP grant (#9901530), and a training grant fellowship from Florida State University NSF-RTG (#9602233).

Received for publication July 26, 2001; revised September 20, 2001; accepted for publication September 25, 2001.

Reprint requests: Kenneth H. Roux, PhD, Department of Biological Science, Biology Unit I, Florida State University, Tallahassee, FL 32306-4370.

0091-6749/2002 $35.00 + 0 **1/87/120558**
doi:10.1067/mai.2002.120558

The specific aim of this research was to screen the large and small subunits of the English walnut allergen Jug r 1 for linear IgE-binding epitopes. We report the finding of a single immunodominant IgE-binding epitope residing in the large subunit, and we have identified 4 core amino acids necessary for this binding to occur.

METHODS

Human sera

The study was approved by the Human Subjects Review Committee at the University of California, Davis. Sera from 20 walnut-allergic patients with convincing histories of life-threatening systemic allergic reactions to walnuts, positive ImmunoCAP assays (Pharmacia Diagnostics, Columbus, Ohio), and evidence of IgE against rJug r 1 (see below) were used in this investigation. Control sera were obtained from atopic patients with no history of walnut sensitivity and 1 patient who had life-threatening reactions to walnuts, positive ImmunoCAP assay, and positive serum IgE immunoblot to walnut proteins but was negative for binding to the 2S albumin.

Expression and isolation of rJug r 1

Two protocols for isolation of rJug r 1 were used. In the first, used to produce rJug r 1 for screening patient sera by Western blotting, the Jug r 1 insert was subcloned into the pPROEXHT expression vector (Gibco BRL, Invitrogen Life Technologies, Carlsbad, Calif) and transformed into *Escherichia coli* BL21-CodonPlus (Stratagene Inc, La Jolla, Calif). Cells were grown overnight at 37°C in 5 mL 2X YT broth with 50 μg/mL ampicillin and chloramphenicol. This was then added to 500 mL 2X YT broth and grown to an A_{600} of 1.0. Isopropyl thiogalactose was added to 0.5 mmol/L, and the culture was grown at 37°C for 4 hours. Cells were pelleted, washed with 0.02 mol/L PBS (pH 7.3), and resuspended in BRL Lysis Buffer (Gibco BRL). Lysozyme was added to 500 μg/mL on ice for 15 minutes; this was followed by the addition of sarkosyl to 1% w/v. The cells were sonicated, and the insoluble matter was pelleted by centrifugation at 12,000*g* for 20 minutes at 4°C. Ni-NTA resin (Gibco BRL) was added in BRL Buffer A (Gibco BRL) to the supernatant and allowed to incubate overnight at 4°C. The resin was washed 4 times with BRL Buffer A, and the His-tagged fusion protein was then eluted with BRL Buffer C (Gibco BRL) containing 0.25% sarkosyl.

In the second protocol, rJug r 1 was expressed in fusion with glutathione-S-transferase and purified after the Frangioni and Neel[30] protocol with slight modifications, as previously described by Teuber et al.[6] This glutathione-S-transferase–rJug r 1 fusion protein was used in all dot-blot assays, as described below.

Immunoblotting of rJug r 1

rJug r 1 samples were either subjected to SDS-PAGE and electrophoretically transferred to nitrocellulose (NC) or blotted directly onto the membrane. For SDS-PAGE analysis of rJug r 1, samples were boiled for 5 minutes in sample buffer (60 mmol/L Tris-HCl, pH 6.8, 2% SDS, 10% v/v glycerol, 0.01% w/v bromophenol blue) and electrophoresis was carried out with an 8-mA constant current through use of a SE600 Vertical Slab Gel Unit (Pharmacia Biotech, Piscataway, NJ). An SDS-PAGE gel (13% monomer acrylamide concentration) with 1 μg of protein per 4 mm was used for immunoblotting, as previously described.[6] Protein was transferred to a 0.22-μm NC membrane (MSI, Westborough, Mass) overnight at 30 V through use of a TE42 Transphor Electro-Transfer Unit (Pharmacia Biotech). The blot was cut into 4-mm-wide strips and blocked for 1 hour at room temperature (RT) in PBS/3% nonfat dry milk/0.2% Triton X-100 (TX-100). Diluted sera, 1:5 v:v in the blocking buffer, were added to the strips and incubated overnight at RT. The strips were then washed for 20 minutes 3 times in PBS/0.01% TX-100 and incubated overnight at RT with equine polyclonal ^{125}I-antihuman IgE (Hycor Biomedical Inc, Garden Grove, Calif) diluted 1:5 in the nonfat milk buffer. The strips were washed for 20 minutes 3 times and exposed to Kodak Biomax x-ray film (Kodak, Rochester, NY) at −70°C for 48 hours. All incubations were carried out with rocking.

For dot-blot assays, a graphite pencil was used to circumscribe 4 mm × 8 mm elipses on a dry 0.4-μm NC membrane (Shleicher & Schuell, Keen, NH). The NC was incubated in distilled-deionized water on a rocking table for 5 minutes and allowed to air-dry before each dot was loaded with 0.11 μg of rJug r 1 in 2 μL distilled-deionized water. The antigen-loaded NC was then rinsed in distilled-deionized water as described above, placed protein-side–up on 3MM paper (Whatman Intl Ltd, Maidstone, United Kingdom), and dried under a 60-W light until all moisture was removed. Once dry, dots were rinsed for 2 minutes in TBS-T (Tris-buffered saline solution [TBS; 20 mmol/L Tris, 137 mmol/L NaCl, pH 7.6] containing 0.2 % Tween-20), blocked for 1 hour at RT in TBS-T containing 2% BSA (Sigma, St Louis, Mo), and then incubated overnight at 4°C with sera diluted 1:40 in TBS-T. Membranes were then washed once for 15 minutes and 3 times for 5 minutes in TBS-T before being incubated for 1 hour at RT with horseradish peroxidase–labeled goat-antihuman IgE (Biosource International, Camarillo, Calif) diluted 1:2000 in TBS-T. Washing in TBS-T was repeated as described above, and the reactive dots were identified after (1) a 5-minute incubation in Enhanced Chemiluminescence Plus (Amersham Pharmacia, Piscataway, NJ), prepared according to the manufacturer's instructions, and (2) subsequent exposure to Kodak X-OMAT x-ray film.

Solid-phase peptide (SPOTs) synthesis

On the basis of the published amino acid sequence of Jug r 1 and our unpublished data determining the subunit cleavage sites,[6] twenty-five overlapping 13–amino acid peptides, offset by 3 amino acids, were synthesized; these corresponded to the entire 64–amino acid length of the large subunit (Fig 1, *A*) and the 26–amino acid length of the small subunit (Fig 1, *B*). Additional peptides bearing the target epitope (QGLRGEEMEEMV), mutated forms of this peptide, and a non–IgE-binding peptide (LSQRGLQSSSV) were also synthesized. Thirteen versions of the target peptide were created through a single-site alanine substitution at each position along the amino acid sequence, and 6 mutated peptides were synthesized through use of multiple alanine substitutions (Table I).

Peptides were synthesized on derivatized cellulose sheets through use of 9-fluorenlymethoxy carbonyl-derived (Fmoc) amino acids, as described by the manufacturer (Genosys Biotechnologies, Inc, The Woodlands, Tex). Briefly, cellulose membranes containing free hydroxy groups were esterfied with an Fmoc–amino acid dissolved in 1-methyl-2 pyrrolidinone. The coupling reaction was followed by washing in N,N-dimethylformamide (DMF), and the amino acids were acetylated with acetic anhydride. The membranes were deprotected by washing in DMF and incubation with a solution of 20% piperidine in DMF; this was followed by another wash in DMF. Coupling, acetylation, and deprotection steps were repeated for each cycle. During the final cycle, a mixture of dichloromethane, trifluoroacetic acid, and triisobutylsilane (1:1:0.05) was used to deprotect the acid-labile amino acid side chains.

IgE binding to solid-phase synthetic peptides (SPOTs analysis)

The peptide-containing membranes were washed in TBS and incubated overnight at RT in blocking solution as directed by the

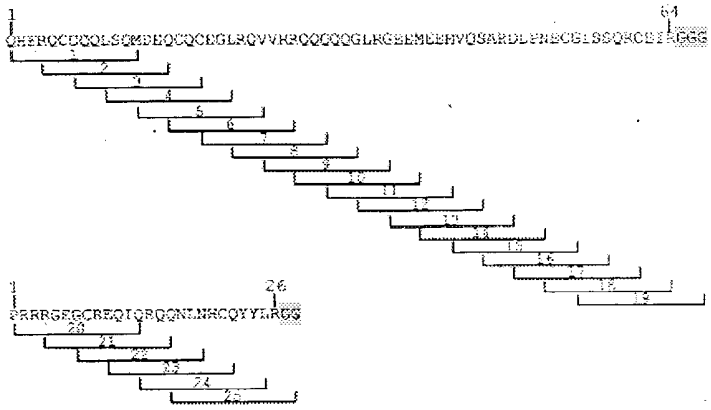


FIG 1. Overlapping peptide synthesis of Jug r 1. **A,** Nineteen 13–amino acid peptides, offset by 3 amino acids, were synthesized; they corresponded to the entire 64–amino acid large subunit. **B,** Similarly, 6 peptides were synthesized; they corresponded to the entire 26–amino acid small subunit. Glycine residues were added to the C-terminal end of each sequence *(shaded)* so that all synthesized peptides were of equal length.

manufacturer (Genosys Biotechnologies, Inc). Membranes were then washed in TBS-T for 10 minutes and incubated overnight at 4°C with an individual patient's serum or pooled patients' sera diluted 1:5 in blocking buffer (total sera:blocking buffer). This incubation was followed by three 5-minute washes in TBS-T and an overnight (4°C) incubation with ^{125}I-antihuman IgE (Hycor Biomedical Inc, Garden Grove, Calif) diluted 1:10 in a mixture of PBS, 5% nonfat dry milk, and 0.05% Tween-20. Three final 10-minute washes in PBS/0.05% Tween-20 were performed, and IgE-peptide reactivity was identified after a 48-hour exposure at –70°C to Kodak Biomax x-ray film.

Soluble epitope peptide synthesis

A soluble form of an identified IgE-reactive peptide (QGLRGEEMEEMV) was synthesized by Fmoc protocols on an automated peptide synthesizer (Model 433A, Applied Biosystems, Foster City, Calif).

Isolation and inhibition of epitope-specific IgE

The epitope-reactive IgE was isolated from patients' sera by affinity chromatography. A 0.5-mL pool of patients' sera, known to be reactive to the reactive synthetic solid-phase epitope peptide, was slowly added to a 5-mL disposable polypropylene column (Pierce Chemical Company, Rockford, Ill) containing cyanogen-bromide-activated beads (Sigma) to which 5 mg of the reactive peptide had been covalently coupled, as described by the manufacturer; the effluent was collected. Bound IgE was eluted with the addition of 0.2 mol/L glycine sulfate (pH 2.3) collected in a beaker containing 10 μL 1% BSA in buffered saline solution borate and subsequently neutralized with 1.0 mol/L Tris. The column was then rinsed with buffered saline solution borate and the initial effluent repassed over the column. The process was repeated for a total of 3 times, yielding 3 eluates and the column effluent.

To test the specificity and reactivity of the epitope-specific IgE, control and pooled and fractionated patients' sera were preincubated overnight at 4°C with different amounts (70, 7.0, and 0.7 μg) of the soluble epitope peptide or a soluble nonspecific peptide. The preincubated sera were used to probe either solid-phase IgE-reactive and non–IgE-reactive peptides or rJug r 1. IgE binding to the solid-phase synthetic peptides and dot-blotted rJug r 1 was detected through use of ^{125}I-antihuman IgE and Enhanced Chemiluminescence Plus, respectively, as described earlier.

TABLE I. Mutational analysis of the Jug r 1 IgE binding epitope, E1

Amino acid sequence*	Substitution	IgE binding†
AGLRGEEMEEMV	Q1A	++
QALRGEEMEEMV	G2A	++
QGARGEEMEEMV	L3A	+
QGLAGEEMEEMV	R4A	–
QGLRAEEMEEMV	G5A	–
QGLRGAEMEEMV	E6A	–
QGLRGEAMEEMV	E7A	–
QGLRGEEAEEMV	M8A	++
QGLRGEEMAEMV	E9A	++
QGLRGEEMEAMV	E10A	+
QGLRGEEMEEAV	M11A	++
QGLRGEEMEEMA	V12A	++
QAARGEEMEEMV	G2A,L3A	+
QAAAGEEMEEMV	G2A,L3A,R4A	–
QAAAAEEMEEMV	G2A,L3A,R4A,G5A	–
QGLRGEEMEAAV	E10A,M11A	+
QGLRGEEMAAAV	E9A,E10A,M11A	–
QGLRGEEAAAAV	M8A,E9A,E10A,M11A	–

*The critical core amino acid residues, located at positions 36 through 39 of the large subunit, and an influential glutamic acid residue at position 42 are indicated in bold type. Mutated residues are underlined.
†Peptides were probed with pooled with pool 1, comprising sera from 6 walnut-sensitive patients.

RESULTS

IgE binding to rJug r 1

IgE from 20 walnut-sensitive patients was shown to bind the 20-kDa His-tagged rJug r 1 (Fig 2). Three sera showed only faint binding, whereas the other 17 sera gave moderate or strong signals. Recombinant Jug r 1

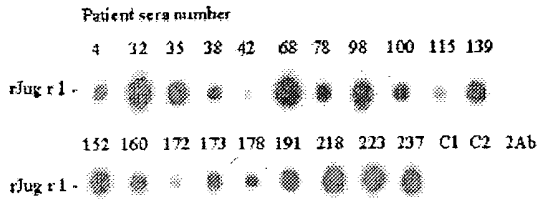
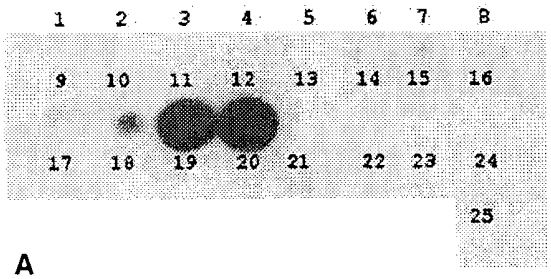


FIG 2. IgE immunoblotting of rJug r 1. *C1*, Serum from a walnut-allergic patient that does not bind to Jug r 1; *C2*, serum from an atopic control patient not allergic to walnuts; *2Ab*, the secondary antibody alone.



A

B

SPOT	Amino Acid Sequence
10	RQQQCQQLRGEEM
11	QQQGLRGEEMEEM
12	GLRGEEMEEMVQS

FIG 3. IgE binding analysis of Jug r 1. **A**, SPOTs array depicts all 25 overlapping 13–amino acid Jug r 1 large and small subunit peptides probed with pool 1, comprising sera from 6 walnut-sensitive patients. **B**, Sequence comparison of 3 adjacent overlapping peptides—10, 11, and 12—recognized by walnut-allergic sera. The amino acid sequence common to each is indicated *(shading)*.

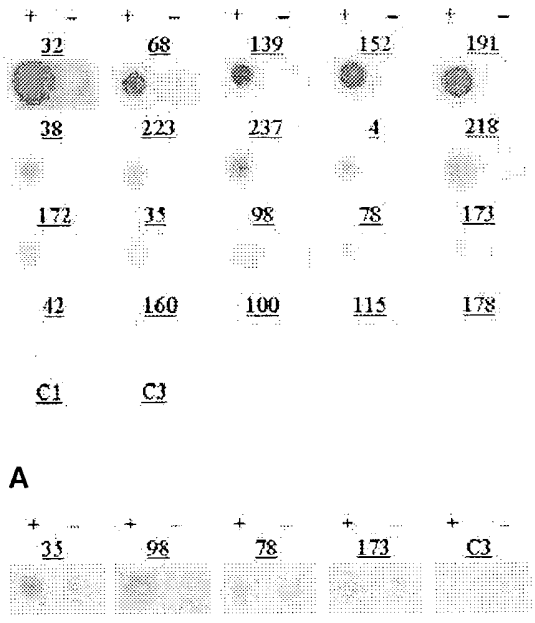


FIG 4. Differential binding of IgE from the sera of 20 walnut-allergic patients to the E1 epitope *(+, left SPOT)* and negative control *(–, right SPOT)* peptides. **A**, Autoradiograph shows patients' sera binding to the E1 peptide strongly (nos. 32, 68, 139, 152, and 191), moderately (nos. 38, 223, 237, 4, 218, and 172), weakly (nos. 35, 98, 78, and 173), or not at all (nos. 42, 160, 100, 115, and 178). **B**, Longer exposure of weakly reactive sera (nos. 35, 98, 78, and 173). *C1*, Serum from a walnut-allergic patient that does not react with Jug r 1; *C3*, serum from an atopic subject not allergic to walnut.

was previously shown to completely inhibit binding to the native 2S albumin and to substantially inhibit binding of patient sera IgE to English walnut in walnut ImmunoCAP assays (up to 84% of measurable antiwalnut IgE)[6]; the folding thus appears to preserve most of the presumed conformational epitopes present on the native mature protein.

Identification and recognition of IgE-reactive linear epitopes on rJug r 1

The entire lengths of both the large and the small subunits of Jug r 1 were studied by probing overlapping solid-phase synthetic peptides with sera from 20 patients randomly assigned to 4 pools. Each pool recognized 3 adjacent peptides from the large subunit; 2 peptides were recognized very strongly (no. 11 and no. 12) and 1 peptide less so (no. 10); no peptides were identified from the small subunit (Fig 3, *A*). A common sequence, GLRGEEM, was observed in all 3 large-subunit peptides (Fig 3, *B*). A fourth partially overlapping sequence (no. 9 in Fig 3, A) showed a slight positive reaction. Some other peptides showed slight positive signals in some assays (eg, no. 7 in Fig 3, *A*) but were not reproducible, leading us to examine only the identified dominant linear epitope-bearing peptides. Additional peptides were tested in which alanine was substituted at each of the 12 amino residues of the IgE-reactive peptide, QGLRGEEMEEMV. In addition, peptides were tested with varying numbers of alanines substituted at the N-terminus and the C-terminus. Together, these data demonstrated that the core amino acids RGEE, at positions 36 to 39, and an additional glutamic acid residue at position 42, were necessary for maximum IgE binding to occur (Table I).

SPOTs containing a 12–amino acid peptide (QGLRGEEMEEMV) bearing the immunodominant epitope, designated *E1*, and SPOTs containing a negative control (–) peptide (LSQRSQQQCRQ) selected from the large subunit of Jug r 1 were used to test the degree of individual reactivity and specificity of allergic and control sera. Of the 20 sera tested, 5 exhibited strong recognition of the epitope peptide, 6 exhibited moderate recognition, 4 exhibited weak recognition, and 5 showed no recognition; none reacted with the (–) peptide (Fig 4).

Isolation and inhibition of E1-specific IgE

To determine whether the identified epitope E1 is the major epitope recognized by patients' sera, we separated E1-specific IgE from the total antiserum and tested both fractions for reactivity with rJug r 1. Epitope-specific IgE

was isolated from patient serum by passage over an E1-affinity column. To assure removal of anti-E1 antibodies, the 3 sequential eluate and the serum effluent fractions were first assayed against our solid-phase IgE-binding (+) and non–IgE-binding (–) peptides. Fig 5, *A* shows binding of IgE to the positive (E1) peptide in unfractionated patient serum, progressively less binding in the first 2 eluate fractions, and no binding in the third fraction, indicating complete removal of the E1-reactive antibody by the column. Significantly, the E1-adsorbed serum (effluent) also showed no reactivity to the E1 peptide, confirming removal of E1-specific IgE from the serum. The eluate and effluent fractions were then used in dot-blot assays to determine whether IgE antibodies to the E1 epitope represent the bulk of the rJug r 1–reactive IgE. As expected, there was demonstration of binding of IgE to rJug r 1 from total sera and from the first 2 eluate (anti-E1) fractions but not from the third (anti-E1-depleted) fraction (Fig 5, *B*). However, there was considerable rJug r 1–specific IgE remaining in the effluent (anti-E1 depleted fraction), demonstrating the presence of additional IgE antibody specific for 1 or more conformational epitopes.

To further examine the specificity of both our peptide-reactive fractions and unfractionated Jug r 1–reactive IgE, a soluble form of the epitope peptide was preincubated with whole patient serum as well as the E1-specific IgE antibodies in an attempt to inhibit their reaction with both our solid-phase peptides and rJug r 1. Varying amounts (70, 7.0, 0.7 μg) of the peptide were incubated with whole patients' sera before the E1 epitope (positive) and negative solid-phase peptides were probed. Partial to complete inhibition was achieved (Fig 5, *C*), demonstrating that the solid- and fluid-phase versions of the peptide are similarly recognized.

The amount of soluble peptide (70 μg) needed to completely inhibit binding of E1-specific IgE to the solid-phase version of this epitope was similarly used for inhibition studies involving unfractionated patient serum IgE and affinity column purified IgE fractions against the rJug r 1. Inhibition was again observed in the eluate (E1 epitope–specific) IgE fractions. These data demonstrate that the epitope is similarly recognized in both the peptide and the recombinant protein. However, IgE reactivity in both whole patient serum and the E1-adsorbed effluent fraction reacted with rJug r 1 (Fig 5, *D*), again demonstrating the existence of additional (presumably conformational) IgE-reactive epitopes on rJug r 1.

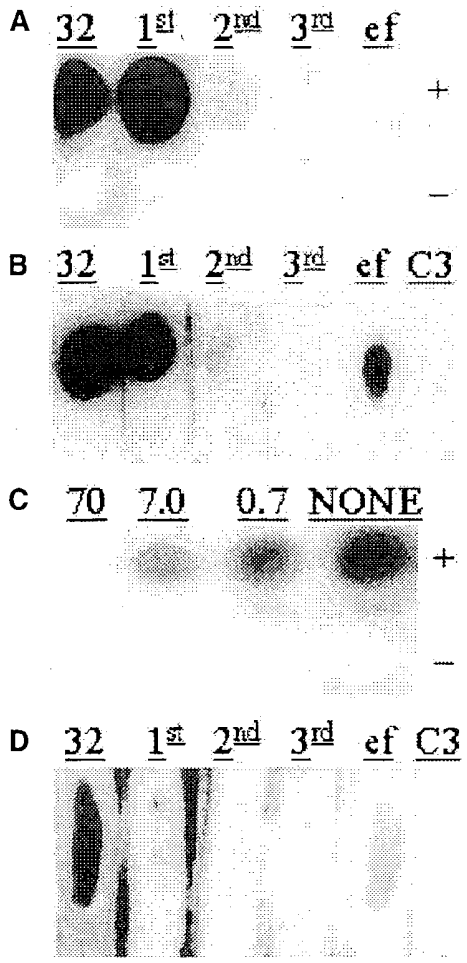


FIG 5. Isolation and inhibition of E1-specific IgE. **A,** Strips containing the E1 epitope peptide *(top SPOT, +)* and a negative control peptide *(bottom SPOT, –)* probed with whole patient serum *(32)*, each of 3 E1 eluates *(1st, 2nd, 3rd)* and the effluent *(ef)*. **B,** Strips dotted with rJug r 1 and probed as in Fig 5, *A*. *C3,* Sera from an atopic subject not allergic to walnut. **C,** Strips containing SPOTs as in Fig 5, *A* and probed with whole patient serum *(32)* containing varied amounts (in μg) of soluble E1 peptide. **D,** Strips dotted with rJug r 1 and probed as in Fig 5, *B*, but with added soluble E1 inhibitor.

DISCUSSION

On a global basis, walnuts rank second in tree nut production.[31] In children, anaphylactic reactions to peanut and tree nuts are responsible for most fatal and near-fatal food-allergic reactions.[3,32] In a recent study, 32 fatal cases of anaphylactic reactions to foods reported to a national registry were analyzed; it was determined that up to 31% were due to tree nuts.[33] The severity and prevalence of some food-induced allergic reactions have prompted investigation to identify and characterize offending allergens.

Jug r 1, a major allergen in the English walnut, has previously been characterized as a 2S albumin seed storage protein.[6] 2S albumins are found in almost all edible seeds. In view of the importance of the members of this class of protein as food allergens and the possibility of cross-reactivity among class-related proteins,[34] a detailed knowledge of these proteins would be valuable. In particular, given the critical role that allergen-specific IgE plays in the allergic reaction, determination of allergen-specific IgE binding epitopes—be they linear or conformational—appears to be of great importance for a better understanding of the allergenic nature of foods and for possible therapeutic intervention.

We have found a single linear IgE-reactive epitope and have defined its core amino acid residues. To date, no common structural character of linear IgE epitopes has been identified,[35] but this could change as more epitope mapping studies are completed. A key step in the allergic reaction is the binding of at least 2 IgE antibodies to a multivalent allergen. The fact that we have found only 1 linear epitope on Jug r 1 is unique in that all previously analyzed allergens have contained multiple linear IgE-binding sites.[35] Although we cannot rule out the existence of weakly binding peptides that are undetected by our methods, the affinity of such peptides would be much lower, as indicated by the high intensity of the signal on the epitope E1. Our findings, in combination with the knowledge that allergens must be multivalent to elicit an allergic response, point toward the existence of conformational epitopes on Jug r 1 as well. Although the data are qualitative, it is noteworthy that all 20 patients' sera bound rJug r 1 (Fig 2), whereas only 15 sera bound the linear epitope, many of these exhibiting disproportionately faint binding. This too is evidence that conformational epitopes are relevant and key in IgE binding for some patients' sera. In fact, it has been estimated that most protein epitopes are conformational.[36] It is worthwhile to note that the labeling of an epitope as "linear," based on the use of synthetic epitopes, is imprecise because the identified epitope might be a fragmented part of a larger, discontinuous epitope.[35]

Clinically, insight into the IgE-binding epitopes of allergens, linear or conformational, is extremely useful. Identifying the immunodominant linear IgE-binding epitope of Jug r 1 might lead to better designs for walnut allergy therapy. Currently, there is no generally agreed-on treatment for IgE-mediated food allergy, and complete avoidance of the food is therefore recommended.[37] With tree nuts such as walnuts, however, avoidance is often difficult because equipment used for food processing is often shared; the implicated allergen thus might be present in trace amounts or might be added to foods in which its presence is not expected by the consumer. Because tree nut allergies are rarely resolved, reactions can be severe, and accidental ingestion is almost inevitable, life for afflicted patients can be fraught with anxiety.[38]

Among new approaches to allergen immunotherapy and diagnosis are use of blocking peptides to inhibit the triggering of IgE-mediated hypersensitivity reactions, immunization with specific peptides to alter the type of immune response elicited, allergen reengineering to enhance protective responses, and use of defined peptides in diagnostic assays.[9-12] A common factor for the application of these approaches is a required knowledge of the amino acids in the IgE-reactive sites of the allergen. We have mapped the major linear IgE epitope of Jug r 1 and identified the critical core amino acid residues within the epitope. Additional studies are planned to identify the conformational epitopes on this allergen.

We thank Rich Peterson for his valuable assistance.

REFERENCES

1. Sampson HA. Food allergy. Part 1: immunopathogenesis and clinical disorders [review]. J Allergy Clin Immunol 1999;103:717-28.
2. Yocum MW, Khan DA. Assessment of patients who have experienced anaphylaxis: a 3-year survey. Mayo Clin Proc 1994;69:16-23.
3. Sampson HA, Mendelson LM, Rosen JP. Fatal and near-fatal anaphylactic reactions to food in children and adolescents. N Engl J Med 1992;327:380-4.
4. Yunginger JW, Sweeney KG, Sturner WQ, Giannanadra LA, Teigland JD, Bray M, et al. Fatal food-induced anaphylaxis. JAMA 1988;260:1450-2.
5. Breiteneder H, Ebner C. Molecular and biochemical classification of plant-derived food allergens. J Allergy Clin Immunol 2000;106:27-36.
6. Teuber SS, Dandekar AM, Peterson WR, Sellers CL. Cloning and sequencing of a gene encoding a 2S albumin seed storage protein precursor from English walnut *(Juglans regia)*, a major food allergen. J Allergy Clin Immunol 1998;101:807-14.
7. Bock SA, Atkins FM. The natural history of peanut allergy. J Allergy Clin Immunol 1989;83:900-4.
8. Sicherer SH, Munoz-Furlong A, Burks WA, Sampson HA. Prevalence of peanut and tree nut allergy in the US determined by a random digit dial telephone survey. J Allergy Clin Immunol 1999;103:559-62.
9. Valenta R, Vrtala S, Focke-Tejkl M, Bugajska-Schretter A, Ball T, Twardosz A, et al. Genetically engineered and synthetic allergen derivatives: candidates for vaccination against Type I allergy. Biol Chem 1999;380:815-24.
10. van Neerven RJJ. The role of allergen-specific T cells in the allergic immune response: relevance to allergy vaccination. Allergy 1999;54:552-61.
11. Bindslev-Jensen C, Poulsen LK. In vitro diagnostic methods in the evaluation of food hypersensitivity. In: Food allergy: adverse reactions to foods and food additives. Boston: Blackwell Science Limited: 1997. p. 137-50.
12. Vrtala S, Ball T, Spitzauer S, Pandjaitan B, Suphioglu C, Knox B, et al. Immunization with purified natural and recombinant allergens induces mouse IgG1 antibodies that recognize similar epitopes as human IgE and inhibit the human IgE-allergen interaction and allergen-induced basophil degranulation. J Immunol. 1998;160:6137-44.
13. Zhao Y, Chait BT. Protein epitope mapping by mass spectrometry. Anal Chem 1994;66:3723-6.
14. Dolecek C, Vrtala S, Laeffer P, Steinberger P, Kraft D, Sheiner O, et al. Molecular characterization of Phl p 2, a major timothy grass (*Phleum pratense*) pollen allergen. FEBS Lett 1993;335:299-304.
15. Ghosh B, Perry MP, Marsh DG. Cloning the cDNA encoding the Amb t V allergen from giant ragweed (*Ambrosia trifida*). Gene 1991;101:231-8.
16. Scheiner O, Kraft D. Basic and practical aspects of recombinant allergens. Allergy 1995;50:384-91.
17. Elsayed S, Titlestad K, Apold J, Aas K. A synthetic hexadecapeptide derived from allergen M imposing allergenic and antigenic reactivity. Scand J Immunol 1980;12:171-5.
18. Atassi H, Atassi MZ. Antibody recognition of ragweed allergen Ra3: localization of the full profile of the continuous antigenic sites by synthetic overlapping peptides representing the entire protein chain. Eur J Immunol 1986;16:229-35.
19. Mazur G, Baur X, Modrow S, Becker WM. A common epitope on major allergens from non-biting midges (Chironomidae). Mol Immunol 1988;25:1005-10.
20. van 't Hof W, van Milligen FJ, van den Berg M, Lombardero M, Chapman MD, Aalberse RC. Epitope mapping of the cat (*Felis domesticus*) major allergen Fel d I by overlapping synthetic peptides and monoclonal antibodies against native and denatured Fel d I. Allergy 1993;48:255-63.
21. Helm RM, Cockrell G, Connaughton C, West CM, Herman E, Sampson HA, et al. Mutational analysis of the IgE-binding epitopes of P34/Gly m Bd 30K. J Allergy Clin Immunol 2000;105:378-84.
22. Helm R, Cockrell G, Herman E, Burks A, Sampson H, Bannon G. Cellular and molecular characterization of a major soybean allergen. Int Arch Allergy Immunol 1998;117:29-37.
23. Banerjee B, Greenberger PA, Fink JN, Kurup VP. Conformational and linear B-cell epitopes of Asp f 2, a major allergen of Aspergillus fumigatus, bind differently to immunoglobulin E antibody in the sera of allergic bronchopulmonary aspergillosis patients. Infect Immun 1999;67:2284-91.
24. van 't Hof W, Driedijk PC, van den Berg M, Beck-Sickinger AG, Jung G, Aalberse RC. Epitope mapping of the Dermatophagoides pteronyssinus house dust mite major allergen Der p II using overlapping synthetic peptides. Mol Immunol 1991;28:1225-32.
25. Walsh BJ, Howden ME. A method for the detection of IgE binding

sequences of allergens based on a modification of epitope mapping. J Immunol Methods 1989;121:275-80.
26. Elsayed S, Holen E, Dybendal T. Synthetic allergenic epitopes from the amino-terminal regions of the major allergens of hazel and birch pollen. Int Arch Allergy Appl Immunol 1989;89:410-5.
27. Burks AW, Shin D, Cockrell G, Stanley JS, Helm RM, Bannon GA. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. Eur J Biochem 1997;245:334-9.
28. Stanley JS, King N, Burks AW, Huang SK, Sampson HA, Cockrell G, Helm RM, West CM, Bannon GA. Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. Arch Biochem Biophys 1997;342:244-53.
29. Rabjohn P, Helm EM, Stanley JS, West CM, Sampson HA, Burks AW, et al. Molecular cloning and epitope analysis of the peanut allergen Ara h 3. J Clin Invest 1999;103:535-42.
30. Frangioni JV, Neel BG. Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. Anal Biochem 1993;210:179-87.
31. Sze-Tao KWC, Sathe SK. Walnuts (Juglans regia L): proximate composition, protein solubility, protein amino acid composition and protein in vitro digestibility. J Sci Food Agric 2000;80:1393-401.
32. Foucard T, Malmheden Yman I. A study on severe food reactions in Sweden—is soy protein an underestimated cause of food anaphylaxis? Allergy 1999;54:261-5.
33. Bock SA, Munoz-Furlong A, Sampson HA. Fatalities due to anaphylactic reactions to foods. J Allergy Clin Immunol 2001;107:191-3.
34. Pastorello EA, Farioli L, Pravettoni V, Ispano M, Conti A, Ansaloni R, et al. Sensitization to the major allergen of Brazil nut is correlated with the clinical expression of allergy. J Allergy Clin Immunol 1998;102:1021-7.
35. Bredehorst R, David K. What establishes a protein as an allergen? [review]. J Chromatogr B Biomed Sci Appl 2001;756:33-40.
36. Berzofsky JA. Intrinsic and extrinsic factors in protein antigenic structure. Science 1985;229:932-40.
37. Smith LJ, Muñoz-Furlong A. The management of food allergy. In: Metcalfe DD, Sampson HA, Simon RA, editors. Food allergy: adverse reactions to foods and food additives. Boston: Blackwell Science Limited; 1997. p. 431-44.
38. Sicherer SH, Sampson HA. Peanut and tree nut allergy. Curr Opin Pediatr 2000;12:567-73.

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

Mutational analysis of the IgE-binding epitopes of P34/Gly m Bd 30K

Ricki M. Helm, PhD,[a] Gael Cockrell, BS,[a] Cathie Connaughton, BS,[a] C. Michael West, BS,[b] Eliot Herman, PhD,[c] Hugh A. Sampson, MD,[d] Gary A. Bannon, PhD,[b] and A. Wesley Burks, MD[a] *Little Rock, Ark, Beltsville, Md, and New York, NY*

Background: Peanuts and soybeans are 2 foods that have been shown to be responsible for many atopic disorders. Because of their nutritional benefit, soybean proteins are now being used increasingly in a number of food products. Previous studies have documented multiple allergens in soybean extracts, including glycinin, β-conglycinin, and the P34/Gly m Bd 30K protein.

Objective: Our overall goal was to identify soybean-specific allergens to begin to understand molecular and immunochemical characteristics of legume proteins. The specific aim of the current investigation was to identify the essential amino acid residues necessary for IgE binding in the 5 distinct immunodominant epitopes of P34/Gly m Bd 30K.

Methods: Serum IgE from 6 clinically sensitive soybean-allergic individuals was used to identify P34/Gly m Bd 30K in the native and single amino acid substituted peptides with use of the SPOTS peptide synthesis technique to determine critical amino acids required for IgE binding.

Results: The intensity of IgE binding and epitope recognition by serum IgE from the individuals varied substantially. With use of serum from 6 clinically soybean-sensitive individuals, 2 of the 5 immunodominant epitopes could be mutagenized to non-IgE binding peptides.

Conclusions: Single-site amino acid substitution of the 5 immunodominant epitopes of Gly m Bd 30K with alanine revealed that IgE binding could be reduced or eliminated in epitopes 6 and 16 in the serum obtained from 6 soybean-sensitive patients. (J Allergy Clin Immunol 2000;105:378-84.)

Key words: Soybean allergen, P34/Gly m Bd 30K, B-cell epitopes

Soybeans and peanuts are both members of the legume plant family and they share common antigenic fractions. Patients allergic to one member often have IgE antibodies that react to other members of the legume family. Although cross-reacting proteins among soybeans and peanuts are immunochemically frequent, clinical responses to more than one is not common.[1-3] A comparison of IgE immunoblots to soybean extracts with common legumes showed that soybean extracts shared several common as well as unique proteins with other members of the legume family.[4,5]

Abbreviations used

DBPCFC:	Double-blind, placebo-controlled food challenge
Gly m Bd:	Glycine max band
rP34:	Recombinant P34 protein
TBST:	TRIS-buffered saline solution with 0.05% Tween 20

Previous studies have documented multiple allergens in soybean extracts, including glycinin, β-conglycinin, and the Glycine max band (Gly m Bd) 30K protein.[6-11] Ogawa et al[7,8] initially described a soybean band designated Gly m Bd 30K as a 7S soybean protein allergen, assigning it as the major allergenic protein in soybean extracts that bound specific IgE in the serum of 65% of patients with atopic dermatitis in Japan. The first 15 amino acids of Gly m Bd 30K were identical to those of a 34-kd soybean seed storage vacuole protein identified by Kalinski et al.[9,10] P34 is an outlying member of the papain superfamily[9,10] that appears to be a *Pseudomonas* defense protein.[11]

The most efficient way to avoid soybean allergy is strict avoidance of food products containing allergenic proteins. Several methods are being conducted to reduce the allergenicity of allergens, including enzymatic methods, physical treatment, or development of soybean seeds that lack the allergenic protein.[12] A combination of both genetically selected soybean cultivars and physicochemical separation methods appears to be the most promising method. With this in mind, our goal was to identify soybean allergens and specifically determine the presence of unique soybean protein epitopes in their amino acid structure that could potentially be mutagenized to non-IgE-binding proteins. The prevalence of the hypersensitivity reactions to the allergen identified as Gly m Bd 30K and its homology to the previously identified protein P34, a constituent of protein storage vacuoles in soybean seed cotyledons, made this protein an ideal candidate for identification and characterization of the major IgE-binding epitopes.[13] The specific aim of the current investigation was to identify the essential amino

From the [a]Department of Pediatrics, University of Arkansas for Medical Sciences, Arkansas Children's Nutrition Center, Little Rock, Ark, the [b]Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, the [c]Climate Stress Laboratory, United States Department of Agriculture, Agriculture Research Service, Beltsville, Md, and the [d]Division of Pediatric Allergy and Immunology, Mt Sinai Medical School, New York, NY.

Supported by a grant from the Arkansas Children's Nutrition Center (No. USDA-ARS 0524-0022[a]) and grant No. AI-24439 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health.[d]

Received for publication July 6, 1999; revised Oct 6, 1999; accepted for publication Oct 13, 1999.

Reprint requests: Ricki M. Helm, PhD, Department of Pediatrics, University of Arkansas for Medical Sciences, Arkansas Children's Nutrition Center 1120 Marshall St, Little Rock, AR 72202.

1/1/103611

acid residues necessary for IgE binding in the 5 distinct immunodominant epitopes in this major soybean allergen. In this article we report the localization of specific amino acids that can be substituted to render immunodominant epitopes 6 and 16 to non-IgE-binding peptides.

MATERIAL AND METHODS

Serum IgE

Twelve patients with atopic dermatitis and a positive immediate skin test response to soybean extracts were confirmed to be sensitive to soybean protein by double-blind, placebo-controlled food challenge (DBPCFC).[2] The individual sera were diluted 1:20 (vol/vol) in TRIS-buffered saline solution (25 mmol/L TRIS–hydrochloric acid, pH 7.2, 150 mmol/L sodium chloride) with 0.05% Tween 20 (TBST) for immunoblot analysis and in 1:5 (vol/vol) TBST with blocking solution as recommended by the manufacturer (Genosys Biotechnologies, The Woodlands, Tex) for mapping of IgE-binding regions and epitopes of the Gly m Bd 30K allergen. The Human Advisory Committee approved all studies at the University of Arkansas for Medical Sciences and Mt Sinai School of Medicine.

Soybean extracts

Preparations of defatted soybean flakes were obtained from commercial sources (Dr Jim Brooks, Ross Laboratories, Columbus, Ohio). Defatted (Soxhlet extraction with petroleum ether) soybean flakes were ground to a fine powder in a mortar and pestle and extracted (1:20 wt/vol) in 20 mmol of sodium phosphate, 1 mmol of sodium chloride, pH 7.2, overnight at 4°C.

A clone designated P34,[9] representing the protein seed storage vacuole protein, was used as a source of recombinant protein. PCR techniques were applied to isolate the protein-coding region of this clone. This portion was then subcloned into the pet-24(b)+ vector for expression of recombinant protein in *Escherichia coli.* The protein was expressed with a 6x-his tag on its carboxyl terminus. The 6x-his tag allowed the purification of full-length protein by nickel resin column affinity chromatography. The purified protein was characterized by immunoblot analysis with serum IgE from soybean-sensitive individuals. Results indicate that the recombinant P34 protein was recognized with a similar IgE-binding intensity as the native protein, suggesting that the recombinant protein can be used for future studies. The protein was desalted into PBS, concentrated (4.2 mg/mL), filter sterilized, and stored for future use.

SDS-PAGE/Western immunoblot analysis

SDS-PAGE was performed according to the method of Laemmli.[14] All gels were composed of 12.5% acrylamide resolving gel and 4% acrylamide stacking gel. Electrophoretic transfer of proteins and immunoblotting were performed according to standard procedures.[15] The blots were blocked with TBST with 1% BSA 60 minutes before incubation with antibodies diluted in TRIS-buffered saline buffer for at least 12 hours at 4°C or for 4 hours at room temperature. Primary IgE antibody was detected with radiolabeled anti-IgE antibody (Sanofi Pasteur Institute, Chaska, Minn) diluted 1:5 (vol/vol) in TRIS-buffered saline buffer. Blots were washed 3 times in buffer, air dried, and exposed to Kodak X-Omat x-ray film at –70°C for 18 to 144 hours and developed in an automatic developer (Konica SRX-101).

IgE binding of mutagenized immunodominant epitopes

The immunodominant epitopes of P34 recognized by 4 clinically soybean-sensitive individuals[13] were used to generate individual peptides for IgE-binding studies. Substitution of an alanine residue for each of the wild-type amino acids in the sequence was used to determine which amino acids were essential for IgE binding. Individual peptides were synthesized on a cellulose membrane containing free hydroxyl groups with use of Fmoc amino acids according to the manufacturer's instructions (Genosys Biotechnologies). Synthesis of each peptide was started by esterification of an Fmoc amino acid to the cellulose membrane. After washing, all residual amino functions on the sheet were blocked by acetylation to render the cellulose membrane unreactive during subsequent steps. Each additional Fmoc amino acid was esterified to the previous one by this same process. After addition of the last amino acid in the peptide chain, the amino acid side chains were deprotected with a mixture of dichloromethane–trifluoroacetic acid–tri-isobutylsilane (1:1:0.05, vol/vol), followed by treatment with dichloromethane and washing with methanol. Membranes containing synthesized peptides were washed with TRIS-buffered saline solution and then incubated with the Genosys blocking solution overnight at room temperature. The membranes were probed immediately with serum IgE as described above or stored at –20°C until used. IgE binding was determined by incubating cellulose membranes containing the synthesized peptides with serum from individuals with documented soybean hypersensitivity and, as control, serum from an individual with high-titer IgE with no known food allergy. Primary antibody was detected with radiolabeled anti-IgE as described for autoradiographs of SDS-PAGE/Western IgE immunoblot analysis.

RESULTS

SDS-PAGE of soybean extracts and recombinant P34

Fig 1 shows a representative profile of an immunoblot of IgE binding to a crude soybean extract and to recombinant P34 (rP34) (lanes 2 and 4, respectively). The IgE-binding is most intense at the 30-kd rP34 protein band. The IgE immunoblot sensitivity reveals radiographic binding at both higher and lower molecular weight bands that are not discernible in the Coomassie blue–stained photograph. These bands are evident in the original stained gel and represent complexes and breakdown products of the rP34 molecule common to many recombinant proteins isolated under these conditions. Control serum from nonleguminous-sensitive patients did not bind to any of the bands from our rP34 material nor did serum IgE from soybean-sensitive individuals bind to extracts of *E coli* in these regions (data not shown).

IgE binding to immunodominant epitopes

The prevalence of serum IgE binding to epitopes 1, 6, 13, 15, and 16 were previously identified as immunodominant epitopes with use of serum from 4 clinically soybean-sensitive individuals.[13] In the current investigation we were able to add data from an additional 8 soybean-sensitive individuals. To determine the extent of patients' serum IgE binding to the immunodominant epitopes, each epitope was synthesized as a single spot on SPOTs membranes and incubated with serum from individual patients. The immunodominant epitopes are listed in Fig 2 with a representative blot of 6 patients' serum IgE-binding pattern. The intensity of IgE binding and epitope recognition by serum IgE from the individuals varied substantially. For example, serum IgE from

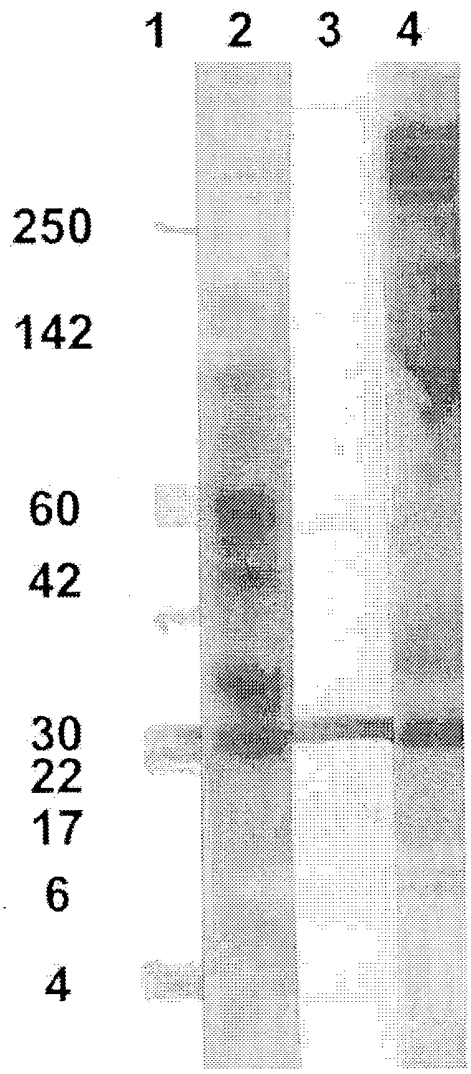


FIG 1. SDS-PAGE IgE immunoblot analysis of crude soybean extract and rP34. *Lane 1,* Molecular weight markers. *Lane 2,* IgE immunoblot autoradiograph of crude soybean extract. *Lane 3,* Coomassie blue–stained rP34 protein. *Lane 4,* IgE immunoblot autoradiograph of rP34 protein.

patients 5 and 3 bound all 5 epitopes with the same intensity. Patients 7, 8, and 1 demonstrated more intense staining to epitope 15, whereas serum from patient 3 recognized epitope 13 as the predominant epitope. This difference in intensity and binding was characteristic of serum IgE from all 12 patients tested. Because the intensity of the binding varied, we chose to test patient serum against the 5 immunodominant epitopes. Table I shows a summary of the 12 patients' serum IgE that were tested for binding to the 5 immunodominant epitopes. We had 2 patients who did not recognize any of the 5 immunodominant epitopes and 5 who recognized all the epitopes, and the remaining bound from 2 to 4 of the 5

TABLE I. Summary of P34 immunodominant epitope mapping by serum IgE from soybean-sensitive individuals

	P34 epitope binding				
	1	6	13	15	16
1*	+	±	+	+	+
2	+	+	±	+	±
3*	+	±	+	+	+
4	+	+	–	+	+
5*	+	+	+	+	+
6*	+	±	+	+	+
7*	+	+	+	+	+
8*	+	±	±	+	+
9	–	+	+	–	+
10	–	–	–	–	–
11	–	–	–	–	–
12	–	+	+	–	+

+, Positive binding; ±, marginal binding; –, negative binding.
*Selected for mutagenized epitope mapping.

TABLE II. Mutagenesis of P34/Gly m Bd 30K epitope 1

SPOT No.	Mutagenized epitope 1	Position change
1	ALVLLLFSLL	F1A
2	FAVLLLFSLL	L2A
3	FLALLLFSLL	V3A
4	FLVALLFSLL	L4A
5	FLVLALFSLL	L5A
6	FLVLLAFSLL	L6A
7	FLVLLLASLL	F7A
8	FLVLLLFALL	S8A
9	FLVLLLFSAL	L9A
10	FLVLLLFSLA	L10A
11	FLVLLLFSLL	Native

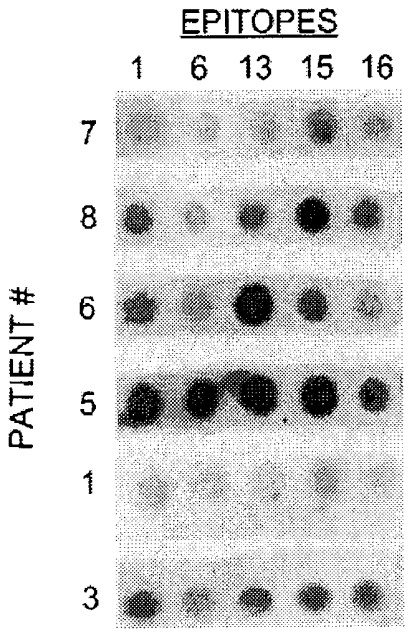
Note: If an alanine residue resided in mutagenized location, glycine was used as substituted amino acid.

immunodominant epitopes. This is in agreement with our initial results in that there appears to be a great amount of diversity in the immunodominant IgE-binding epitopes.

IgE binding to mutagenized immunodominant epitopes

To determine the amino acids that are critical to IgE binding of the 5 immunodominant epitopes, we mutagenized the amino acid sequences by substituting alanine at each position and compared the IgE binding to native (wild-type) epitope. Table II depicts the position change for epitope 1. Each epitope was modified in this manner in a series of replicates for analysis with IgE from 6 of the P34-sensitive individuals. Fig 3 shows the results from 3 of these individuals.

Serum IgE from patient 12 recognized only epitopes 6, 13, and 16. Alanine substitutions in epitope 6 at positions 6 and 7 and at position 7 in epitope 16 showed reduced IgE binding. Alanine substitutions at positions 1-6 and 8 showed increased and position 7 reduced IgE binding with this patient's serum IgE to epitope 16. Epitope 13



P34/Gly m Bd 30K Immunodominant Epitopes

Epitope	Amino acid position	Amino acid sequence
1	3 - 12	FLVLLLFSLL
6	110 - 119	PQEFSKKTYQ
13	229 - 238	RCKANKIQDK
15	299 - 308	INHFVLLVGY
16	311 - 340	GYIWIQRNTG

FIG 2. Autoradiograph of patient serum IgE binding to 5 immunodominant P34/Gly m Bd 30K epitopes.

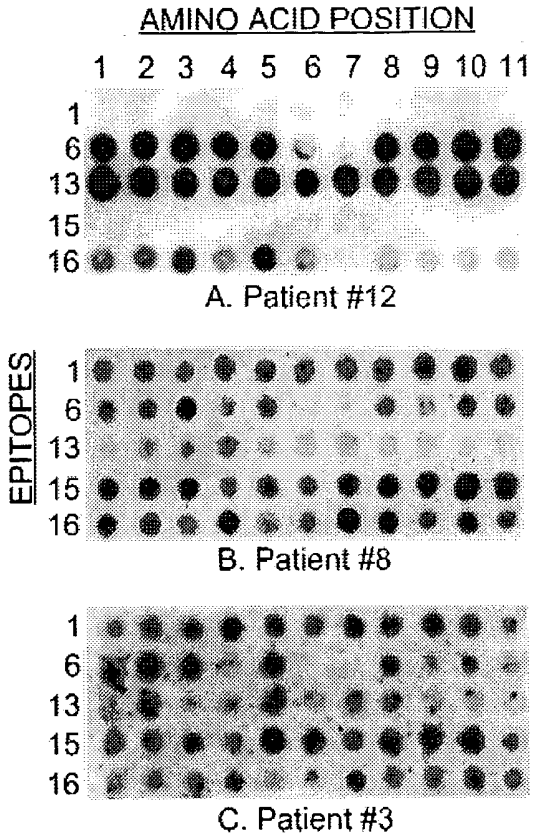


FIG 3. Autoradiograph of representative patient serum IgE binding to mutagenized P34/Gly m Bd 30K epitopes.

could not be mutagenized to a non-IgE-binding peptide with alanine substitutions at any position in the peptide with serum from patient 12.

Serum IgE from patients 8 and 3 recognized all 5 epitopes. Consistent with patient 12, serum from both patients failed to bind IgE at positions 6 and 7 of epitope 6. Serum from patient 8 demonstrated that an alanine substitution did not significantly alter the IgE binding in the remaining sites of the epitopes investigated. The autoradiograph for patient 3, in addition to reduced IgE binding at position 6 and 7 of epitope 6, revealed both increased and decreased levels of radiostained intensity in the remaining sites of the immunodominant epitopes.

Table III summarizes the extent of IgE binding to mutagenized epitopes with the 6 individuals we tested. Notable in this table is the consistency with which we could or could not mutagenize IgE binding by alanine substitution. Epitopes 1, 13, and 15 could not be mutagenized to a non-IgE-binding peptide, whereas epitope 6 could be altered to a non-IgE-binding peptide by an alanine substitution at positions 6 and 7 in the amino acid sequence. Epitope 16 was more variable and could be mutagenized to a non-IgE-binding peptide at position 7 for 4 patients (patients 4, 5, 7, and 12) and at positions 5 and 6 for 2 patients (patients 3 and 8). This suggested that we could mutagenize epitopes 6 and 16 to non-IgE-binding peptides by substituting alanine at specific sites in the sequence of rP34 for patients 3, 8, and 12.

DISCUSSION

Five immunodominant IgE-binding epitopes were previously identified with use of serum IgE from 4 soybean-sensitive individuals by B-cell epitope mapping with overlapping peptides for the soybean protein identified as Gly m Bd 30K (Gly m Bd30K/P34), a plant thiol protease.[13] This protein has been identified as the major soybean allergen with a frequency of 65% in patients with atopic dermatitis with food allergy.[8] In this study we report the results of mutagenizing the 5-immunodominant epitopes on the P34/Gly m Bd 30K soybean for the identification of amino acids critical for IgE binding. The single-site amino acid substitution of the 5 epitopes revealed that IgE binding could be reduced or eliminated depending on the serum of individual patients.

Adverse reactions to various food products are a com-

TABLE III. Summary of 6 patients' sera used to map mutagenized immunodominant epitopes

	Amino acid position (alanine)										
Epitope	1	2	3	4	5	6	7	8	9	10	Weight
Patient 12											
1	–	–	–	–	–	–	–	–	–	–	–
6	+	+	+	+	+	–	–	+	+	+	+
13	+	+	+	+	+	+	+	+	+	+	+
15	–	–	–	–	–	–	–	–	–	–	–
16	+	+	+	+	+	+	–	+	+	+	+
Patient 9											
1	–	–	–	–	–	–	–	–	–	–	–
6	+	+	+	+	+	–	–	+	+	+	+
13	+	+	+	+	+	+	+	+	+	+	+
15	–	–	–	–	–	–	–	–	–	–	–
16	+	+	+	+	+	+	–	+	+	+	+
Patient 4											
1	+	+	+	+	+	+	+	+	+	+	+
6	+	+	+	+	+	–	–	+	+	+	+
13	–	–	–	–	–	–	–	–	–	–	–
15	+	+	+	+	+	+	+	+	+	+	+
16	+	+	+	+	+	+	–	+	+	+	+
Patient 3											
1	+	+	+	+	+	+	+	+	+	+	+
6	+	+	+	+	+	–	–	+	+	+	+
13	+	+	+	+	+	+	+	+	+	+	+
15	+	+	+	+	+	+	+	+	+	+	+
16	+	+	+	+	–	–	+	+	+	+	+
Patient 5											
1	+	+	+	+	+	+	+	+	+	+	+
6	+	+	+	+	+	–	–	+	+	+	+
13	+	+	+	+	+	+	+	+	+	+	+
15	+	+	+	+	+	+	+	+	+	+	+
16	+	+	+	+	+	+	–	+	+	+	+
Patient 8											
1	+	+	+	+	+	+	+	+	+	+	+
6	+	+	+	+	+	–	–	+	+	+	+
13	+	+	+	+	+	+	+	+	+	+	+
15	+	+	+	+	+	+	+	+	+	+	+
16	+	+	+	+	–	–	+	+	+	+	+

mon problem faced by primary care physicians. Among the numerous reactions ascribed to soy products, hypersensitivity, in contrast to enterocolitis, is described as an abnormal immunologic response in which IgE is the principal immunoglobulin. In most studies the diagnosis of soy allergy has been based on clinical evaluation or case histories reported by parents without substantial scientific diagnostic criteria. In a recent study 1% of children with allergic disease had documented soy allergy and all of them had atopic dermatitis.[16] In related studies with a 3- to 5-year follow-up, the frequency of soy reactions was almost nonexistent,[17] with other studies proclaiming a more transient adverse soy reaction.[18,19] The reports suggest a wide discrepancy in soy reactions, ranging from nonexistent to transient and documented soy reactions. This discrepancy is manifested in patient selection criteria and the method used to document soy reactions. Our study involved patients with soy reactions clinically proved by DBPCFC. After diagnosis of a food allergy, the only proved therapy is elimination of the offending allergen.[20] For those individuals with soybean allergy a more direct approach to treatment of this food allergy is desirable.

Although the incidence of allergy to soybean proteins is quite low in comparison with other major food proteins, the gradually increasing consumption of soybean products makes the identification and characterization of major soy allergens a focus for investigation. Herian et al[5] identified 3 different patterns of immunoblots in adult patients with soybean allergy on the basis of molecular weights. Serum bound IgE patterns were resolved into those individuals who had IgE-recognizing proteins with molecular weights of 50 to 60 kd, 20 kd, or 14 kd. Awazuhara et al[21] identified 5 proteins with molecular weights ranging from 20 to 78 kd in the whey and globulin fraction of soybean extracts. None of these proteins had an IgE-binding frequency greater than 50%. In studies with 8 serum samples from soy protein–allergic children, immunoblots to crude soy proteins demonstrat-

ed that IgE- and IgG-specific antibodies to crude soy extracts were elevated in patients with positive DBPCFC, but none of the soy fractions investigated were clearly more antigenic.[6] In a study carried out in Japan, SDS-PAGE and immunoblotting analyses revealed that 19% of patients with atopic dermatitis possessed IgE antibodies specific for soybean proteins.[8] At least 16 proteins were recognized, with detection frequencies ranging from 1.4% to 2.9% for 2S and 11S components to as high as 65% for the 7S-globulin fraction that contains Gly m Bd 30K. The remaining protein components had IgE binding frequencies of less than 30%. The variability in the number of epitopes recognized reflects both the polyclonal nature of the immune response and the subject variability in epitope recognition.

The development of methods to reduce the allergenicity of soybean proteins is required because no other food material is used in as many foods as soybean throughout the world. Hidden soybean allergens in processed foods have been shown to provoke severe reactions in sensitive individuals.[22] A combination of both genetically selected soybean cultivar mutants lacking the α- and α′-subunits of conglycinin and physicochemical separation methods to remove Gly m Bd 30K from soy milk has been described.[12] Although this is a very simple method, efficient industrial application procedures have not been developed to produce a nonallergenic or very-low-allergenic soybean protein source.

Another method to reduce the allergenicity of soybean proteins would be to introduce genes that would produce hypoallergenic versions of the allergens.[23-25] The intensity of binding to wild-type peanut allergen was routinely high, whereas binding to mutated epitopes was markedly decreased at specific sites in each of the peanut epitopes. In contrast, the IgE binding to wild-type P34/Gly m Bd 30K epitopes did not show the same degree of intense binding as evidenced by autoradiographic analysis, nor could we mutate each of the epitopes to non-IgE-binding peptides. Peanut sensitivity, which is a life-long, life-threatening hypersensitivity, is a more severe food allergy leading to anaphylactic episodes. Soybean hypersensitivity, which is usually outgrown by age 5 years or less, is associated more with gastrointestinal symptoms and normally does not lead to an anaphylactic episode.

The elucidation of the IgE-binding epitopes on P34/Gly m Bd 30K, the major allergen of soybean in atopic dermatitis identified by Ogawa et al[8] and the determination of amino acids essential to IgE binding could provide the information necessary to alter the gene responsible for encoding a non-IgE-binding protein. Although other proteins are involved in soybean hypersensitivity, we chose the P34 molecule as our first soybean allergen to mutagenize and to test for insertion into the soybean genome. This approach will require that we suppress the native gene and determine that our altered gene product will retain its native function and properties in the soybean plant and seed. We propose to test our hypothesis with use of in vitro transformation techniques to determine the potential of producing a soybean product with reduced P34 serum IgE binding. Our overall approach is to prepare moderately modified proteins by site-directed amino acid substitutions of soybean allergens and use of cosuppression of native allergens to produce a hypoallergenic soybean plant.

REFERENCES

1. Sampson HA. Immunopathogenic role of food hypersensitivity in atopic dermatitis. Acta Derm Venereol 1992;176:34-7.
2. Burks AW, Mallory SB, Williams LW, Shirrell MA. Atopic dermatitis: clinical relevance of food hypersensitivity reactions. J Pediatr 1988;113:447-51.
3. Bernhisel-Broadbent J, Sampson HA. Cross allergenicity in the legume botanical family in children with food hypersensitivity. J Allergy Clin Immunol 1989;83:435-40.
4. Burks AW, Williams LW, Helm RM, Connaughton C, Cockrell G, O'Brien T. Identification of a major peanut allergen Ara h I, in patients with atopic dermatitis and positive peanut challenge. J Allergy Clin Immunol 1991;88:172-9.
5. Herian AM, Taylor SL, Bush RK. Identification of soybean allergens by immunoblotting with sera from soy-allergic adults. Int Arch Allergy Appl Immunol 1990;92:193-8.
6. Burks AW, Brooks JR, Sampson HA. Allergenicity of major component proteins of soybean determined by enzyme-linked immunosorbent assay (ELISA) and immunoblotting in children with atopic dermatitis and positive soy challenges. J Allergy Clin Immunol 1988;81:1135-42.
7. Ogawa T, Bando N, Tsuji H, Okijima H, Nishikawa L, Sasoka K. Investigations of the IgE-binding proteins in soybeans by immunoblotting with sera from soybean-sensitive patients with atopic dermatitis. J Nutr Sci Vitamin 1991;37:555-65.
8. Ogawa R, Tsuji H, Bando N, Kitamura K, Zhu YL, Hirano H, et al. Identification of the soybean allergenic protein, Gly m Bd 30k, with the soybean seed 34-kDa oil-body-associated protein. Biosci Biotech Biochem 1993;57:1030-3.
9. Kalinski A, Weisemann JM, Matthews BF, Herman EM. Molecular cloning of a protein associated with soybean seed oil bodies that is similar to thiol proteases of the papain family. J Biol Chem 1990;265:13843-48.
10. Kalinski A, Melroy DL, Dwivedi RS, Herman EM. A soybean vacuolar protein (P34) related to thiol proteases is synthesized as a glycoprotein precursor during seed maturation. J Biol Chem 1992;267:12068-76.
11. Ji C, Boyd C, Slaymaker D, Okinaka U, Herman EM, Keen NT. Purification and characterization of a 34 kDa syringolide binding protein from soybean. Proc Natl Acad Sci U S A 1998;95:3306-11.
12. Samoto M. Takahaski K, Fukuda Y, Nakamura S, Kawamura Y. Substantially complete removal of the 34 kDa allergenic soybean protein, Gly m Bd 30K from soy milk of a mutant lacking the α and α′-subunits of conglycinin. Biosci Biotech Biochem 1996;60:1911-3.
13. Helm RM, Cockrell G, Herman E, Burks AW, Sampson HA, Bannon GA. Cellular and molecular characterization of a major soybean allergen. Int Arch Allergy Immunol 1998;117:29-37.
14. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680-5.
15. Towbin H, Staehlin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedures and some applications. Proc Natl Acad Sci U S A 1979;76:4350-4.
16. Bruno G, Giampietro PG, Del Guerico MJ, Gallia P, Giovannini L, Lovati C, et al. Soy allergy in not common in atopic children: a multicenter study. Pediatr Allergy Immunol 1997;8:190-3.
17. Cantani A, Lucenti P. Natural history of soy allergy and/or intolerance in children, and clinical use of soy-protein formulas. Pediatr Allergy Immunol 1997;8:59-74.
18. Hattevig G, Kjellman B, Johansson SGO, Bjorksten B. Clinical symptoms and IgE responses to common food proteins in atopic and healthy children. Clin Allergy 1984;14:551-5.
19. Host A, Halken S. A prospective study of cow milk allergy in Danish infants during the first three years of life. Allergy 1990;45:587-96.
20. Sampson HA. Adverse reactions to foods. In: Middleton E Jr, Reed CE,

Ellis EF, Adkinson NF Jr, Yunginger JW, Busse WW. editors. Allergy, principles and practice. St Louis: Mosby–Year Book; 1993. p 1661-86.

21. Awazuhara H, Kawai H, Maruchi N. Major allergens in soybean and clinical significance of IgG4 antibodies investigated by IgE- and IgG4-immunoblotting with sera from soybean-sensitive patients. Clin Exp Allergy 1997;27:325-32.
22. Vidal C, Perez-Carral C, Chomon B. Unsuspected sources of soybean exposure. Ann Allergy Asthma Immunol 1997;79:350-2.
23. Burks AW, Shin D, Cockrell G, Stanley JS, Helm RM, Bannon GA. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. Eur J Biochem 1997;245:334-9.
24. Stanley JS, King N, Burks AW, Huang SK, Sampson HA, Cockrell G, et al. Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. Arch Biochem Biophys 1997;343:244-53.
25. Rabjohn P, Helm EM, Stanley JS, West CM, Sampson HA, Burks AW, et al. Molecular cloning and epitope analysis of the peanut allergen, Ara h 3. J Clin Invest 1999;103:535-42.

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

Article title	Modulation of IgE-binding properties of tree pollen allergens by site-directed mutagenesis
Article identifier	0065259897014648
Authors	Ferreira_F Rohlfs_A Hoffmann-Sommergruber_K Schenk_S Ebner_C Briza_P Jilek_A Kraft_D Breitenbach_M Scheiner_O
Journal title	Advances in Experimental Medicine and Biology
ISSN	0065-2598
Publisher	Plenum
Year of publication	1996
Volume	409
Issue	1996
Supplement	0
Page range	127-135
Number of pages	9
User name	Adonis
Cost centre	
PCC	$16.00
Date and time	Wednesday, September 22, 2004 10:51:21 AM

MODULATION OF IgE-BINDING PROPERTIES OF TREE POLLEN ALLERGENS BY SITE-DIRECTED MUTAGENESIS

Fátima Ferreira,[1] Angelika Rohlfs,[1] Karin Hoffmann-Sommergruber,[2] Siegfried Schenk,[2] Christof Ebner,[1] Peter Briza,[1] Alexander Jilek,[1] Dietrich Kraft,[2] Michael Breitenbach,[1] and Otto Scheiner[2]

[1]Inst. f. Genetik u. Allg. Biologie
Universität Salzburg
A-5020 Salzburg, Austria
[2]Inst. f. Allg. u. Exp. Pathologie
Universität Wien
A-1090 Vienna, Austria

1.INTRODUCTION

In the last few years, the use of recombinant DNA techniques for the characterization of atopic allergens offered new aspects for diagnosis and therapy of Type I allergies (1). It has made it possible to produce large quantities of well characterized "wild-type" recombinant allergens, or first-generation recombinant products, and many of them are now being developed for diagnosis and possibly therapy of Type I allergies. For instance, recombinant Asp f 1 expressed in *E. coli* was successfully used for serologic and clinical diagnosis of A. fumigatus allergy (2). It has been demonstrated that the use of two recombinant birch pollen allergens, Bet v 1 and Bet v 2 (profilin), allows accurate *in vitro* (ELISA, immunoblots) and *in vivo* (skin prick test, intradermal test) diagnosis of birch pollen allergy (3,4). In addition, recombinant Bet v 1 could be efficiently used for identifying food cross-sensitization induced by Bet v 1-related proteins (4). These studies demonstrate that recombinant allergens are adequate tools for in vivo and in vitro allergen-specific diagnosis, which might be considered as an important step towards allergen-specific therapy. Presently, specific-immunotherapy is performed using natural allergen extracts that may contain, besides the desired allergen, other unwanted components.

In addition to their use for characterization of allergens, recombinant DNA techniques also offered the unique possibility of arbitrarily altering the nucleotide sequence of a gene and, subsequently, the sequence of the encoded protein in order to produce novel "mutant" proteins, or second generation recombinant products, displaying altered proper-

New Horizons in Allergy Immunotherapy
edited by Sehon et al. Plenum Press, New York, 1996

ties. Previously we have shown that isoforms of Cor a 1, the major hazel pollen allergen, displayed striking differences in their ability to bind IgE from allergic patients (5). Since these isoforms showed high amino acid sequence similarity, we speculated that the differences in IgE-binding was a result of sequence dissimilarities. In particular, the exchange at position 10 from a threonine in Bet v 1a and Cor a 1/11 isoform to proline in Cor a 1/16 isoform seemed to correlate to its lower IgE-binding capacity. In this study, we have tested this hypothesis using a PCR-based site-directed mutagenesis approach to produce a single amino acid exchange in the Cor a 1/16 isoallergen.

2. MATERIALS AND METHODS

2.1. DNA Constructs

The cDNAs coding for Bet v 1a (6) and Cor a 1/16 (5) were cloned in the pMW175 expression vector (7). The amino acid exchange at position 10 (Pro→Thr) in Cor a 1/16 was engineered by PCR-mediated mutagenesis using the following primers:

NcoI mutagenic primer 5'-GGGCCATGGGTGTTTTCAATTACGAGGTT-GA-GACCACCTCCGTT-3', base exchanged indicated in bold; NcoI site is underlined.

EcoRI primer 5'-CCCGAATTCTTAGTTGTATTCAGCAGAGTGTGCGAA-3', EcoRI site is underlined.

The PCR was performed with 1 ng template (pMW175/Cor a 1/16 construct) and 1 μM of each primer, using 30 cycles of 1 min denaturation at 95°C, 1 min annealing at 42°C, and 1 min extension at 72°C. The PCR product was digested with NcoI and EcoRI and subcloned in the pMW175 expression vector. The resulting plasmids were used to transform competent *E. coli* BL21 cells. All PCR amplified products were sequenced according to the dideoxi chain termination method (8).

2.2. Expression of Cor a 1/16, Cor a 1/16T10, and Bet v 1a

E. coli BL21 cells containing the pMW175/Bet v 1a, pMW175/Cor a 1/16, and pMW175/Cor a 1/16T10 plasmids were grown and expression of the recombinant allergens induced by adding IPTG to a final concentration of 1 mM. After incubation at 37°C for 6 h, cells were harvested by centrifugation and lysed by repeated freeze-thaw cycles.

2.3. SDS-Page and Immunoblots

E. coli lysates of recombinant Cor a 1/16, Cor a 1/16T10, and Bet v 1a or purified proteins were analyzed by SDS-PAGE according to the method of Laemmli (9), using 15% acrylamide gels. Proteins were visualized by staining with Coomassie Brilliant Blue R-250. For immunoblot analysis, proteins were separated by 15% SDS-PAGE and electroblotted onto nitrocellulose membranes. Immunoblots using a monoclonal anti-Bet v 1, BIP 1, were performed as described previously (10). IgE immunoblots were carried out using sera from birch pollen allergic patients. Bound IgE was detected using ^{125}I-rabbit anti-human IgE. *E. coli* lysates harboring the plasmid without insert were used as a control. In all experiments, reagents, and cell lysates were from identical batches and were used in the same concentrations. Autoradiography was performed at -70°C for 12–48 h with intensifying screens.

2.4. Purification of Recombinant Allergens

rBet v 1a, rCor a 1/16, and rCor a 1/16T10 proteins were purified from crude bacterial lysates by chromatofocusing and reversed-phase HPLC (10). Purified proteins were analysed by SDS-PAGE according to the method of Laemmli (9) and visualized by staining with Coomassie Brilliant Blue R-250.

2.5. Trypsin Treatment, and Matrix-Assisted Laser Desorption/ Ionization Mass Spectrometry (MALDI-MS) of Cor a 1/16 and Cor a 1/16T10 Proteolytic Fragments

Purified Cor a 1 proteins (100 µg in distilled water) were heated for 20 min at 95°C and diluted with an equal volume of 0.2 M NH_4HCO_3. One microgram of trypsin was added and the mixture incubated at 37°C for 2 h. Afterwards, trypsin was added again, and incubation continued for an additional 4-h period. The reaction was stopped by adding 1/10 vol of trifluoracetic acid and dried *in vacuo*. The resulting peptide mixtures were subjected to MALDI-MS analysis using the HP G2025A system equipped with a nitrogen laser.

2.6. T-Cell Proliferation Assays

Isolation of Bet v 1-specific T-cell clones from the peripheral blood of birch pollen allergic patients (as indicated by typical case history, positive skin tests, and positive RAST) and proliferation assays were done as previously described (11).

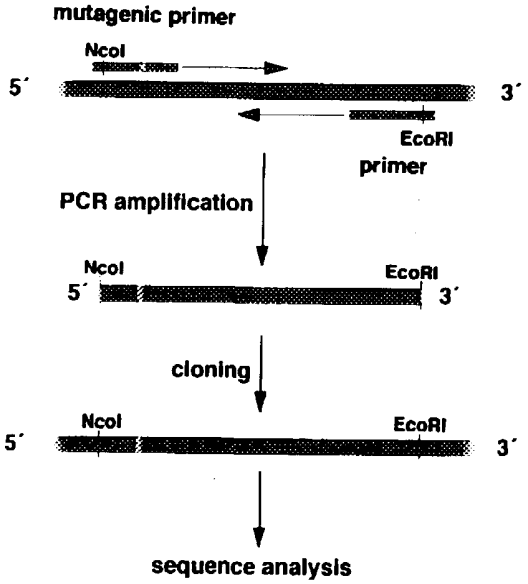


Figure 1. PCR mutagenesis of the Cor a 1/16 cDNA. The 5' Nco mutagenic primer and the 3' EcoRI primer were used in a PCR with the Cor a 1/16 cDNA.

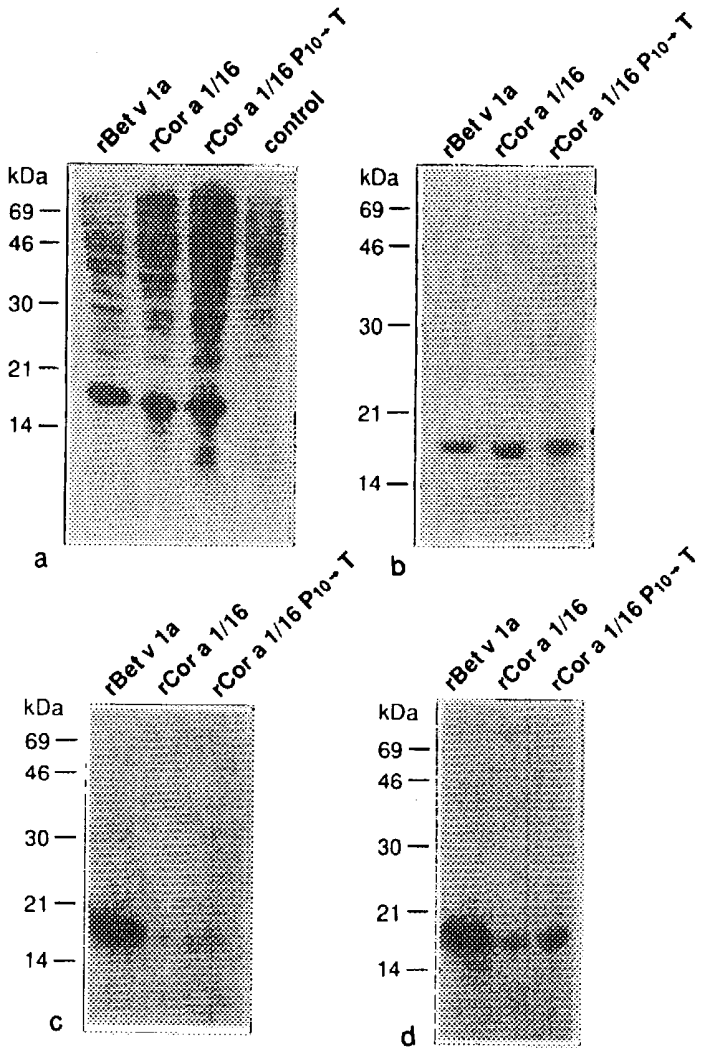


Figure 2. Expression, purification and immunoblot analysis of Cor a 1/16T10 mutant protein. (a), Coomassie-stained 15% SDS-polyacrylamide gel of lysates of *E. coli* BL21 host strain containing Cor a 1/16, Cor a 1/16T10, and Bet v 1a pMW175 expression plasmids. Control, host strain BL21 lysate containing the pMW175 expression vector without an insert. (b), Coomassie-stained 15% SDS-polyacrylamide gel of purified rCor a 1/16, rCor a 1/16T10 and rBet v 1a. (c) and (d), Immunoblots of purified rCor a 1/16, rCor a 1/16T10, and rBet v 1a probed with BIP 1, a monoclonal anti-Bet v 1, and with a polyclonal anti-rBet v 1a, respectively.

3. RESULTS AND DISCUSSION

3.1. Expression, Purification, and Mass Spectrometry Analysis of Cor a 1/16 Proteins

In this study, we have used PCR-mediated mutagenesis to generate a mutant of the Cor a 1/16 allergen. A mutant primer was used that had the proline codon at amino acid position 10 replaced by a threonine codon. The strategy for the construction of this Cor a 1/16 mutant (Cor a 1/16T10) is outlined in Figure 1.

The cDNAs coding for Bet v 1a, Cor a 1/16 and Cor a 1/16T10 were subcloned in the pMW175 expression vector and high-levels of recombinant non-fusion proteins were produced by induction with IPTG. Figure 2A shows a Coomassie-stained gel of expressed Bet v 1a, Cor a 1/16, and Cor a 1/16T10 proteins.

The recombinant proteins were purified from crude bacterial lysates using chromatofocusing and reversed-phase HPLC. The proteins appeared homogeneous as determined by SDS-PAGE and Coomassie-staining (Figure 2B).

In order to confirm at the protein level the sequence of the Cor a 1/16T10 mutant allergen, purified rCor a 1/16 and rCor a 1/16T10 were digested with trypsin and the proteolytic fragments subjected to MALDI-MS. Nine peptides were detected for both Cor a 1/16 and Cor a 1/16T10, and their molecular weights were determined from the obtained spectra (Table 1). The observed mass signals could be easily matched with the molecular weight of peptides predicted from the amino acid sequence deduced from the published Cor a 1/16 sequence (T1, T4-T6, T9–10, T12, T18–19). These peptides covered about 70% of the Cor a 1/16 sequence. All peptides detected by MALDI-MS of rCor a 1/16 digests were also detected in rCor a

Table 1. Mass determination of tryptic fragments T1-T19 of rCor a 1/16T10. Theoretical m/z values give the calculated masses of the peptides plus one proton $[M+H]^+$

Fragment	*Sequence*	*m/z*	
		theor.	*observed*
T1	GVFNYEVETPSVISAAR	1840.03	1839.6
T1 $P_{10} \rightarrow T$	GVFNYEVETTSVISAAR	1844.02	1843.2
T2	LFK	407.53	-
T3	SYVLDGDK	896.96	-
T4	LIPK	470.62	470.3
T5	VAPQAITSVENVGGNGGPGTIK	2067.29	2067.3
T6	NITFGEGSR	981.04	981.2
T7	YK	310.36	-
T8	YVK	409.49	-
T9	ERVDEVDNTNFK	1466.54	1466.6
T10	YSYTVIEGDVLGDKLEK	1929.96	1929.4
T11	VCSELK	678.81	-
T12	IVAAPGGGSTLK	1071.25	1071.2
T13	ISSK	434.50	-
T14	FHAK	502.58	-
T15	GDHEINAEEMK	1273.35	-
T16	GAK	275.32	-
T17	EMAEK	607.69	-
T18	LLR	401.52	401.5
T19	AVETYLLAHSAEYN	1581.71	1581.7

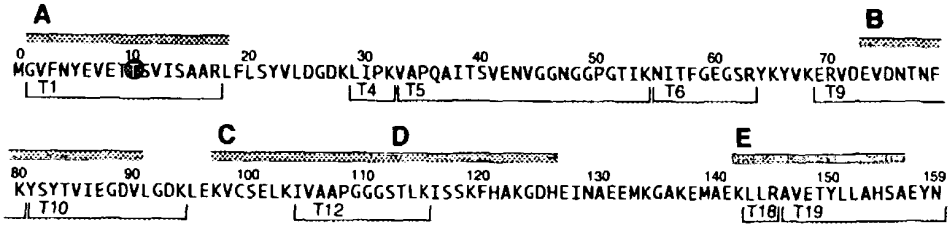


Figure 3. Tryptic peptides of rCor a 1/16T10 (T1-T19) identified by MALDI-MS. The positions of peptides stimulating specific T cell clones (A-E) are indicated by boxes above the sequence.

1/16T10 digests, except for the signal at m/z 1839 corresponding to the N-terminal peptide T1 (Table 1). According to the Cor a 1/16T10 cDNA sequence, peptide T1 should contain the single amino acid exchange (Pro → Thr) when compared to Cor a 1/16. In tryptic digests of rCor a 1/16T10 a signal at m/z 1843 (T-P10 → T) corresponded exactly to the expected mass of T1 with an exchange of proline for a threonine (Table 1). Figure 3 shows the recorded mass signals of rCor a 1/16T10 proteolytic digests mapped onto the cDNA-derived Cor a 1/16T10 sequence according to their molecular mass and enzyme specificity. Thus, the single amino acid substitution engineered in the Cor a 1/16 cDNA to produce Cor a 1/16T10 was also confirmed at the protein level by mass spectrometry analysis of proteolytic digests of rCor a 1/16T10 mutant protein.

3.2 Immunological Properties of Cor a 1/16T10

To evaluate the antibody-binding properties of purified rCor a 1/16T10 mutant protein in comparison to rCor a 1/16 and rBet v 1a, we performed immunoblotting experiments.

Figures 2C and 2D show immunoblots of the purified proteins using a monoclonal anti-Bet v 1 antibody, BIP 1, (Fig. 2C) and a rabbit anti-rBet v 1a serum (Fig. 2D). Both antibodies showed strong reactivity to rBet v 1a. In contrast, BIP 1 did not react with rCor a 1/16 or with rCor a 1/16T10, and the polyclonal anti-Bet v 1a showed a weak reactivity to both rCor a 1/16 and rCor a 1/16T10 proteins.

Immunoblots experiments using sera from birch pollen allergic patients showed remarkable differences in the IgE-binding properties of rCor a 1/16 and rCor a 1/16T10. Figure 4 shows the IgE-binding patterns of rBet v 1a, rCor a 1/16, and rCor a 1/16T10 using sera from five birch pollen allergic patients. All patients showed a marked increase in IgE-binding to rCor a 1/16T10 mutant protein compared to wild type rCor a 1/16, except patient 2. For patients 1, 4, and 5, the replacement of proline at position 10 by a threonine resulted in a change of the antibody-binding pattern from "no-binding" to strong IgE-binding. Interestingly, rCor a 1/16T10 mutant protein in some cases displayed higher IgE-binding activity than rBet v 1a (patients 1, 2, 3, and 5).

Presently, there are no data available on the 3D-structure of Bet v 1 or homologous proteins. Also, there are no precise informations available about IgE-binding motif(s) on the Bet v 1 or Bet v 1-related allergens. However, there are indications that IgE-binding structures on the Bet v 1 molecule might be determined by the protein conformation (10).

Among all standard amino acids, proline seems to occupy a unique position. Proline imposes strong conformational constraints on the peptide chain because the side-chain is cyclized back onto the backbone amide position. When present inside an alpha-helix, the possibility of making hydrogen bonds to the preceding turn is hindered and a kink of 20°

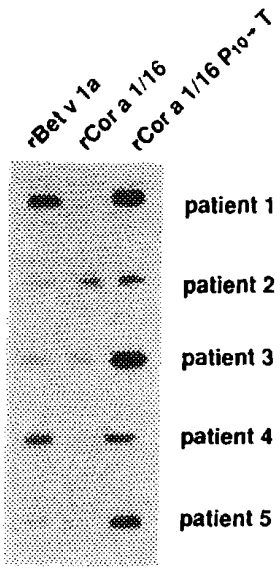


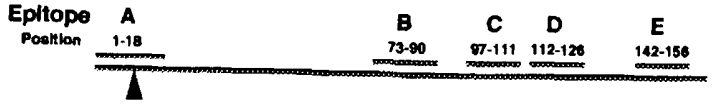
Figure 4. IgE-binding to rCor a 1/16T10, rCor a 1/16, and rBet v 1a. Immunoblot experiments showing serum IgE-reactivity of 5 birch pollen allergic patients.

or more will be introduced in the alpha-helix. In addition, proline can introduce structural heterogeneity since the X-proline (X being any amino acid) bond can assume either the stereoisomeric cis or trans conformation (for a review see 12).

Taking into consideration the exceptional properties of proline, it is possible that the substitution in Cor a 1/16 of proline-10 for a threonine residue might cause conformational changes with the result of a dramatic increase in its IgE-binding activity. This correlates well with the fact that Bet v 1a, which displays high IgE-binding activity, has a threonine at position 10. It will be interesting to test whether the substitution of threonine-10 in Bet v 1a for a proline will lower its IgE-binding activity.

3.3 Activation of Allergen-Specific T Cell Clones

The ability of rCor a 1/16T10 to activate allergen-specific T cell clones was evaluated using Bet v 1a-specific T cell clones. We tested fourteen clones that were established from the peripheral blood of birch pollen allergic patients and were shown to recognize distinct epitopes (A-E) scattered over the whole Bet v 1a molecule (Fig. 5) (ref. 11). Seven of these Bet v 1a-specific T cell clones also recognize the corresponding sequence on Cor a 1/16. The epitope recognized by the clone RR9 comprises the amino acid substitution on rCor a 1/16T10. This clone reacted with both rCor a 1/16 and rCor a 1/16T10 mutant protein. Except for the clone WD25, all other clones reacting with rCor a 1/16 also reacted with rCor a 1/16T10. It is not clear why this clone failed to proliferate in response to rCor a 1/16T10. As shown in Fig. 5, this clone recognizes an epitope (epitope E) corresponding to the C-terminal region (142–156) of Bet v 1a and Cor a 1/16. The possibility of a mutation in the peptide epitope recognized by this clone can be ruled out since MALDI-MS analysis of proteolytic fragments of rCor a 1/16T10 confirmed the structural integrity of the region corresponding to epitope E (see Fig. 3 and Table 1). It is conceiv-



TCC	EPITOPE	rBet v 1a	rCor a 1/16	rCor a 1/16 P_{10}→T
BE6	A	+	-	-
WD27		+	-	-
RR9		+	+	+
MH23	B	+	-	-
WF22		+	++	++
WF20		+	-	-
MS2	C	+	-	-
WF17		+	-	-
TF1		+	-	-
WF38	D	+	+	+
RR4	E	+	+	+
WD24		+	+	+
TF10		+	+	+
WD25		+	+	-

Figure 5. Proliferative responses of human Bet v 1-specific T cell clones (TCC) to rCor a 1/16, rCor a 1/16T10, and rBet v 1a. On top, the black bar represents the Bet v 1a amino acid sequence. The positions of the epitopes (A-E) recognized by the TCC are indicated by boxes above the Bet v 1a sequence.

able that the single amino acid exchange at position 10 in Cor a 1/16T10 can affect the conformation of the protein to an extent that the processing by antigen presenting cells is different and a non-reactive peptide is created or that the epitope is destroyed. In this respect, Finnegan and Amburgey (13) showed that a single amino acid change in the staphylococcal nuclease protein affects the structure of the processed peptides in such a manner that stimulatory determinants are no longer presented to certain T cell clones.

4. CONCLUSIONS

The results presented here suggest that it is possible to modulate the IgE-binding properties of tree pollen allergens by single amino acid substitutions at crucial positions. This finding makes it possible to develop second-generation of recombinant allergens with antibody-binding properties specifically modulated for diagnosis and for therapy.

Following this line, we are presently testing the effect of single amino acid exchanges on the IgE-binding properties of Bet v 1a, which were based on the patterns of amino acid substitutions of Bet v 1 isoforms displaying low IgE-binding properties.

The identification of positions crucial for IgE binding might be facilitated by cloning and sequencing isoforms of a particular allergen and determining their IgE-binding properties. In addition, isoforms with low or no IgE-binding activity could be useful tools for defining IgE-binding structures on allergens.

ACKNOWLEDGMENTS

This work was supported by grants P10019-MOB (to F.F. and M.B.), S06704-MED (to C.E.), and S06707-MED (to O.S.).

5. REFERENCES

1. Scheiner, O., Bohle, B., Breitenbach, M., Breiteneder, H., Duchene, M., Ebner, C., Ferreira, F., Hoffmann-Sommergruber, K., Pettenburger, K., Rumpold, H., Steiner, R., Tejkl, M., Valenta, R., Kraft, D. (1992) Recombinant allergens: production and possible clinical applications. In: Godard P., Bousquet J., Michel F.B.,eds. Advances in Allergology and Clinical Immunology, Carnforth, Park Ridge:Parthenon Publishing Group, pp.115–127.
2. Moser, M., Crameri, r., Menz, G., Schneider, T., Dudler, t., Virchow, C., Gmachl, M., Blaser, K., Suter, M. (1992) Cloning and expression of recombinant Aspergillus fumigatus allergen I/a (rAsp fI/a) with IgE binding and Type I skin test activity. J. Immunol. 149, 454–460.
3. Menz, G., Dolocek, C., Schönheit-Kenn, U., Ferreira, F., Moser, M., Schneider, T., Suter, M., Boltz-Nitulescu, G., Ebner, C., Kraft, D., Valenta, R. Serological and skin test diagnosis of birch pollen allergy with recombinant Bet v 1, the major birch pollen allergen. Clin. Exp. Allergy (in press).
4. Pauli, G., Oster, J.P., Deviller, P., Heiss, S., Bessot, J.C., Susani, M., Ferreira F., Kraft, D., Valenta, R. Skin testing with recombinant allergens rBet v 1 and rBet v 2: Diagnostic value for birch pollen and associated allergies. J. Allergy Clin. Immunol. (in press).
5. Breiteneder, H., Ferreira, F., Hoffmann-Sommergruber, K., Ebner, C., Breitenbach, M., Rumpold, H., Kraft, D., Scheiner, O. (1993) Four recombinant isoforms of Cor a I, the major allergen of hazel pollen, show different IgE-binding properties. Eur. J. Biochem. 212, 355–362.
6. Breiteneder, H., Pettenburger, K., Bito, A., Valenta, R., Kraft, D., Rumpold, H., Scheiner, O., Breitenbach, M. (1989) The gene coding for the major birch pollen allergen Bet v 1, is highly homologous to a pea disease resistance response gene. EMBO J. 8, 1935–1938.
7. Studier, F.W., Rosenberg, A.H., Dunn, J.J., Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 185, 60–89.
8. Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.
9. Laemmli, U.K. (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature 227, 680–685.
10. Ferreira, F., Hoffmann-Sommergruber, K., Breiteneder, H., Pettenburger, K., Ebner, C., Sommergruber, W., Steiner, R., Bohle, B., Sperr, W., Valent, P., Kungl, A., Breitenbach, M., Kraft, D., Scheiner, O. (1993) Purification and characterization of recombinant Bet v 1, the major birch pollen allergen. Immunological equivalence to natural Bet v 1. J. Biol. Chem. 268, 19574–19580.
11. Ebner, C., Szepfalusi, Z., Ferreira, F., Jilek, A., Valenta, R., Parronchi, P., Maggi, E., Romagnani, S., Scheiner, O., Kraft, D. (1993) Identification of multiple T cell epitopes of Bet v 1, the major birch pollen allergen, using allergen-specific T cell clones and overlapping peptides. J. Immunol. 150, 1047–1054.
12. Vanhoof, G., Goossens, F., De Meester, I., Hendriks, D., Scharpe, S. (1995) Proline motifs in peptides and their biological processing. FASEB J. 9, 736–744.
13. Finnegan, A. and Amburgey, C.F. (1989) A single amino acid mutation in a protein antigen abrogates presentation of certain T cell determinants. J. Exp. Med. 170, 2171–2176.

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 342, No. 2, June 15, pp. 244–253, 1997
Article No. BB979998

Identification and Mutational Analysis of the Immunodominant IgE Binding Epitopes of the Major Peanut Allergen *Ara h 2*

J. Steven Stanley,*† Nina King,* A. Wesley Burks,† Shau K. Huang,‡ Hugh Sampson,‡ Gael Cockrell,† Ricki M. Helm,† C. Michael West,† and Gary A. Bannon*,[1]

*Department of Biochemistry & Molecular Biology and †Department of Pediatrics, University of Arkansas for Medical Sciences, Arkansas Children's Hospital Research Institute, Little Rock, Arkansas 72205; and ‡Division of Pediatric Allergy & Immunology, Johns Hopkins University, Baltimore, Maryland 21287

Received December 27, 1996, and in revised form March 14, 1997

A major peanut allergen, *Ara h* 2, is recognized by serum IgE from >90% of patients with peanut hypersensitivity. Biochemical characterization of this allergen indicates that it is a glycoprotein of ~17.5 kDa. Using N-terminal amino acid sequence data from purified *Ara h* 2, oligonucleotide primers were synthesized and used to identify a clone (741 bp) from a peanut cDNA library. This clone was capable of encoding a 17.5-kDa protein with homology to the conglutin family of seed storage proteins. The major linear immunoglobulin E (IgE)-binding epitopes of this allergen were mapped using overlapping peptides synthesized on an activated cellulose membrane and pooled serum IgE from 15 peanut-sensitive patients. Ten IgE-binding epitopes were identified, distributed throughout the length of the *Ara h* 2 protein. Sixty-three percent of the amino acids represented in the epitopes were either polar uncharged or apolar residues. In an effort to determine which, if any, of the 10 epitopes were recognized by the majority of patients with peanut hypersensitivity, each set of 10 peptides was probed individually with serum IgE from 10 different patients. All of the patient sera tested recognized multiple epitopes. Three epitopes (aa27-36, aa57-66, and aa65-74) were recognized by all patients tested. In addition, these three peptides bound more IgE than all the other epitopes combined, indicating that they are the immunodominant epitopes of the *Ara h* 2 protein. Mutational analysis of the *Ara h* 2 epitopes indicate that single amino acid changes result in loss of IgE binding. Two epitopes in region aa57-74 contained the amino acid sequence DPYSP that appears to be necessary for IgE binding. These results may allow for the design of improved diagnostic and therapeutic approaches to peanut hypersensitivity. © 1997 Academic Press

Immediate hypersensitivity reactions to foods occur in 6–8% of children and about 1% of adults (1, 2) and are mediated by the production of immunoglobulin E (IgE)[2] antibodies to glycoproteins present in the food (3). Peanuts are a major cause of serious allergic reactions in both children and adults. The hypersensitivity to peanuts often starts in childhood and continues throughout life. This is in contrast to other childhood food allergies such as to milk and eggs which generally resolve spontaneously with age (4). In addition, peanut allergy is more likely to cause fatal anaphylaxis than any other food allergy (5–7). Currently, avoidance is the only effective means of dealing with food allergy, but the use of peanuts and peanut by-products as supplements in many different foods makes accidental consumption almost inevitable. Thus, the prevalence and chronic nature of peanut allergy, the potential severity of the allergic reaction, and the widespread use of peanuts in consumer foods necessitates improved methods for managing peanut hypersensitivity.

Food hypersensitivity reactions occur shortly after contact of a specific allergen with its corresponding IgE antibodies which are bound to mast cells. Cross-linking of the allergen-specific IgE by the respective allergen activates the mast cells to release histamine, heparin, and other mediators responsible for the clinical symptoms observed. Thus, the IgE-binding epitopes of the allergens play an important role in the disease process.

[1] To whom correspondence should be addressed at UAMS, Slot 516, 4301 W. Markham, Little Rock, AR 72205. Fax: (501) 686-8169.

[2] Abbreviations used: IgE, immunoglobulin E; DMF, *N,N*-dimethylformamide; TBS, Tris-buffered saline.

0003-9861/97 $25.00

If improved diagnostic and therapeutic capabilities are to be developed, it is important to determine the primary structure of the major allergens, the IgE-binding sites of these allergens, and the frequency of recognition of any IgE-binding epitopes. Characterization of allergens will provide a better understanding of the human immune response involved in food hypersensitivity reactions.

Several studies have shown that the most allergenic portion of the peanut is the protein fraction of the cotyledon (8–10). Two highly abundant glycoproteins found in the cotyledon are the peanut allergens, *Ara h* 1 and *Ara h* 2. These proteins are recognized by serum IgE from >90% of peanut-sensitive patients, thus establishing them as important allergens (11, 12). The majority of serum IgE recognition of the *Ara h* 1 and *Ara h* 2 allergens appear to be due to epitopes within these proteins that are linear amino acid sequences that do not contain significant amounts of carbohydrate (13). We have recently described the cloning of the gene encoding the *Ara h* 1 allergen and identified it as a seed storage protein belonging to the vicilin family of legume storage proteins (13).

In this paper we report the cloning and nucleotide sequence of the second major peanut allergen, *Ara h* 2. The derived amino acid sequence has been used to construct synthetic peptides and perform a detailed examination of the linear IgE binding epitopes of this protein. These results may allow for the design of improved diagnostic and therapeutic approaches to peanut hypersensitivity.

MATERIALS AND METHODS

Serum IgE. Serum from 15 patients with documented peanut hypersensitivity (mean age, 25 years) was used to identify peanut allergens. Each of these individuals had a positive immediate skin prick test to peanut and either a positive double-blind, placebo-controlled, food challenge or a convincing history of peanut anaphylaxis (laryngeal edema, severe wheezing, and/or hypotension). Details of the challenge procedure and interpretation have been discussed previously (11). Representative individuals with elevated serum IgE levels (who did not have peanut specific IgE or peanut hypersensitivity) were used as controls in these studies. At least 5 ml of venous blood was drawn from each patient and allowed to clot, and the serum was collected. All studies were approved by the Human Use Advisory Committee at the University of Arkansas for Medical Sciences.

Isolation and amino acid sequence analysis of peanut allergen Ara h 2. *Ara h* 2 was purified to near homogeneity from whole peanut extracts according to the methods of Burks *et al.* (12). Purified *Ara h* 2 was electrophoresed on 12.5% acrylamide mini-gels (Bio-Rad Laboratories, Hercules, CA) in Tris/SDS/glycine buffer. The gels were stained with 0.1% Coomassie blue in 10% acetic acid and 50% methanol and destained in 40% methanol for 3 h with continuous shaking. Gel slices containing *Ara h* 2 were sent to the W. M. Keck Foundation (Biotechnology Resource Laboratory, Yale University, New Haven, CT) for amino acid sequencing. Amino acid sequencing of intact *Ara h* 2 and tryptic peptides of this protein was performed on an Applied Biosystems sequencer with an on-line HPLC column that was eluted with increasing concentrations of acetonitrile.

Peanut RNA isolation and Northern (RNA) gels. Three commercial lots from the 1979 crop of medium grade peanut species, *Arachis hypogaea* (Florunner), were obtained from North Carolina State University for this study. Total RNA was isolated from 1 g of this material according to procedures described by Larsen (14). Poly(A)$^+$ RNA was isolated using a purification kit (Collaborative Research, Bedford, MA) according to manufacturer's instructions. Poly(A)$^+$ RNA was subjected to electrophoresis in 1.2% formaldehyde agarose gels, transferred to nitrocellulose, and hybridized with ^{32}P-labeled probes according to the methods of Bannon *et al.* (15).

Computer analysis of Ara h 2 sequence. Sequence analysis of the *Ara h* 2 gene was done on the University of Arkansas for Medical Science's Vax computer using the Wisconsin DNA analysis software package. The algorithm of Needleman and Wunsch was used to align the complete amino acid sequence of *Ara h* 2 with homologous proteins before determining the percent identity.

cDNA expression library construction and screening. Peanut poly(A)$^+$ RNA was used to synthesize double-stranded cDNA according to the methods of Watson and Jackson (17) and Huynh *et al.* (18). The cDNA was treated with *Eco*RI methylase and then ligated with *Eco*RI and *Xho*I linkers. The DNA was then ligated with *Eco*RI–*Xho*I cut, phosphatase treated λ-ZAP XR phage arms (Stratagene, LaJolla, CA), and *in vitro* packaged. The library was 95% recombinants carrying insert sizes >400 bp. The library was screened using an IgE antibody pool consisting of an equal volume of serum from each patient with peanut hypersensitivity. Detection of primary antibody was with I^{125}-labeled anti-IgE antibody performed according to the manufacturer's instructions (Sanofi, Chaska, MN).

PCR amplification of the Ara h 2 mRNA sequence. Using the oligonucleotide CA(AG)CA(AG)TGGGA(AG)TT(AG)CA(AG)GG(N)-GA(TC)AG derived from amino acid sequence analysis of the *Ara h* 2 peanut allergen as one primer and a 23-nt primer which hybridizes to the Bluescript vector, the cDNA that encodes *Ara h* 2 was amplified from the IgE-positive clones. Reactions were carried out in a buffer containing 3 mM $MgCl_2$, 500 mM KCl, and 100 mM Tris–HCl, pH 9.0. Each cycle of the polymerase chain reaction consisted of 30 s at 95°C, followed by 1 min at 56°C, and 2 min at 72°C. Thirty cycles were performed with both primers present in all cycles. From this reaction, a clone carrying an approximately 700-bp insert was identified.

DNA sequencing and analysis. DNA sequencing was done according to the methods of Sanger *et al.* (19) using either ^{32}P-end labeled oligonucleotide primers or a automated ABI model 377 DNA sequencer using fluorescent tagged nucleotides. Most areas of the clone were sequenced at least twice and in some cases in both directions to ensure an accurate nucleotide sequence for the *Ara h* 2 gene.

Peptide synthesis. Individual peptides were synthesized on a derivatised cellulose membrane using Fmoc amino acid active esters according to the manufacturer's instructions (Genosys Biotechnologies, Woodlands, TX). Fmoc–amino acid derivatives were dissolved in 1-methyl-2-pyrrolidone and loaded on marked spots on the membrane. Coupling reactions were followed by acetylation with a solution of 4% (v/v) acetic anhydride in *N,N*-dimethylformamide (DMF). After acetylation, Fmoc groups were removed by incubation of the membrane in 20% (v/v) piperdine in DMF. The membrane was then stained with bromophenol blue to identify the location of the free amino groups. Cycles of coupling, blocking, and deprotection were repeated until the peptides of the desired length were synthesized. After addition of the last amino acid in the peptide, the amino acid side chains were deprotected using a solution of dichloromethane/trifluoroacetic acid/triisobutlysilane (1/1/0.05). Membranes were either probed immediately or stored at −20°C until needed.

IgE binding assay. Cellulose membranes containing synthesized peptides were washed with Tris-buffered saline (TBS) and then incubated with blocking solution overnight at room temperature. After blocking, the membranes were incubated with serum from patients with peanut hypersensitivity diluted (1:5) in a solution containing TBS and 1% bovine serum albumin for at least 12 h at 4°C or 2 h at

TABLE I

Amino Acid Sequence of *Ara h* 2 Peptides

Peptide	Amino acid sequence
I	X-Q-Q-W-E-L-Q-G-D-R-R-R-Q-S-Q-L-E-R
II	A-N-L-R-P-C-E-Q-H-L-M-Q-K

Note. The amino acid sequence of the amino terminus (I) and a tryptic peptide (II) derived from *Ara h* 2 protein was determined. The sequence is shown as the one-letter amino acid code. X indicates an amino acid that was unable to be assigned.

room temperature. Primary antibody was detected with ^{125}I-labeled anti-IgE antibody (Sanofi).

RESULTS

Isolation and partial amino acid sequence determination of the Ara h 2 protein. The amino terminus of the purified *Ara h* 2 protein, or peptides resulting from trypsin digestion of this protein, were used for amino acid sequence determination. The amino acid sequence representing the amino terminus of the *Ara h* 2 protein (peptide I) and a tryptic peptide fragment (peptide II) is noted in Table I. It was possible to determine the first 17 residues from peptide I and the first 13 residues from peptide II of the major peptide in each fraction. These results confirm and extend previous amino acid sequence analysis of the *Ara h* 2 protein (12).

Identification and characterization of clones that encode peanut allergen Ara h 2. RNA isolated from the Florunner variety of peanuts (*A. hypogaea*) was used to construct an expression library for screening with serum IgE from patients with peanut hypersensitivity. Numerous IgE binding clones were isolated from this library after screening 10^6 clones with serum IgE from a pool of patients with reactivity to most peanut allergens by Western blot analysis. Since the number of plaques reacting with serum IgE was too large to study all in detail, we randomly selected 63 positive clones for further analysis. The inserts from each of these clones were then amplified using vector-specific primers and PCR, separated by agarose gel electrophoresis, and blotted onto nitrocellulose. To identify which of the clones encoded the *Ara h* 2 allergen, a hybridization probe was constructed using a radioactive oligonucleotide [CA(AG)CA(AG)TGGGA(AG)TT(AG)CA(AG)GG(N)-GA(TC)AG] developed from amino acid sequence determined for peptide I and used to probe the amplified inserts. Utilizing this approach, two plaques with ~700-bp inserts were identified. DNA sequence revealed that the selected clones carried identical 741-base inserts which included a 21-base poly(A) tail and a 240 base 3′ noncoding region. This insert contained a large open reading frame starting with an CTC codon

Ara h 2

```
    CTC ACC ATA CTA GTA GCC CTC GCC CTT TTC CTC CTC GCT GCC CAC GCA  48
1     L   T   I   L   V   A   L   A   L   F   L   L   A   A   H   A

    TCT GCG AGG CAG CAG TGG GAA CTC CAA GGA GAC AGA AGA TGC CAG AGC  96
17    S   A   R   Q   Q   W   E   L   Q   G   D   R   R   C   Q   S

    CAG CTC GAG AGG GCG AAC CTG AGG CCC TGC GAG CAA CAT CTC ATG CAG 144
33    Q   L   E   R   A   N   L   R   P   C   E   Q   H   L   M   Q

    AAG ATC CAA CGT GAC GAG GAT TCA TAT GAA CGG GAC CCG TAC AGC CCT 192
49    K   I   Q   R   D   E   D   S   Y   E   R   D   P   Y   S   P

    AGT CAG GAT CCG TAC AGC CCT AGT CCA TAT GAT CGG AGA GGC GCT GGA 240
65    S   Q   D   P   Y   S   P   S   P   Y   D   R   R   G   A   G

    TCC TCT CAG CAC CAA GAG AGG TGT TGC AAT GAG CTG AAC GAG TTT GAG 288
81    S   S   Q   H   Q   E   R   C   C   N   E   L   N   E   F   E

    AAC AAC CAA AGG TGC ATG TGC GAG GCA TTG CAA CAG ATC ATG GAG AAC 336
97    N   N   Q   R   C   M   C   E   A   L   Q   Q   I   M   E   N

    CAG AGC GAT AGG TTG CAG GGG AGG CAA CAG GAG CAA CAG TTC AAG AGG 384
113   Q   S   D   R   L   Q   G   R   Q   Q   E   Q   Q   F   K   R

    GAG CTC AGG AAC TTG CCT CAA CAG TGC GGC CTT AGG GCA CCA CAG CGT 432
129   E   L   R   N   L   P   Q   Q   C   G   L   R   A   P   Q   R

    TGC GAC TTG GAC GTC GAA AGT GGC GGC AGA GAC AGA TAC TAA         474
145   C   D   L   D   V   E   S   G   G   R   D   R   Y   END
```

FIG. 1. Nucleotide sequence Of An *Ara h* 2 cDNA clone. The nucleotide sequence is shown on the first line. The second line is the derived amino acid sequence. Amino acid residues in bold correspond to the determined amino acid sequences of peptide I and II of *Ara h* 2 (Table I). The numbers on the right of the figure indicate the position of the nucleotide sequence relative to the first nucleotide in the inset. The numbers on the left of the figure indicate the position of the amino acid sequence relative to the first amino acid encoded by this clone.

TABLE II

Ara h 2 Sequence Similarities

Protein	Source	% Similar
Conglutin-δ	Lupin	39
Mabinlin I (chain B)	Caper	32–35
2S albumin	Sunflower	34
	Castor bean	30
α-Amylase inhibitor	Wheat	29
CM3 protein	Wheat	27

Note. The *Ara h* 2 nucleotide and derived amino acid sequences were used to search the GenBank, Swiss-Prot, and EMBL databases for any homologous proteins. The table lists the proteins that had the highest similarity to the *Ara h* 2 sequence, the plant source of those proteins, and the percentage similarity between that protein and *Ara h* 2.

and ending with a TAA stop codon at nucleotide position 474 (Fig. 1). The calculated size of the protein encoded by this open reading frame was ~17.5 kDa, which is in good agreement with the molecular weight of *Ara h* 2 that has been determined experimentally (12). With the exception of a single cysteine residue at position 30, the amino acid sequence that was determined from the purified *Ara h* 2 protein (Table I) was found in this clone. The additional coding region on the amino terminal end of this clone probably represents a signal peptide which would be cleaved from the mature *Ara h* 2 allergen.

To determine what size mRNA this clone identified, a ^{32}P-labeled insert was used as a hybridization probe of a Northern blot containing peanut poly(A)$^+$ RNA (data not shown). This insert hybridized to an ~0.7-kb mRNA. The size of the cloned insert and the size of the mRNA were in good agreement. In addition, there was good agreement between the calculated and determined size of the *Ara h* 2 protein. Furthermore, the identity of the determined amino acid sequence from the *Ara h* 2 peptides agreed with that which was determined from the clone. From these data we concluded that an *Ara h* 2 specific clone had been isolated.

Peanut allergen Ara h 2 is a conglutin-like seed storage protein. A search of the GenBank, Swiss-Prot, and EMBL databases revealed significant amino acid sequence homology between the *Ara h* 2 protein and

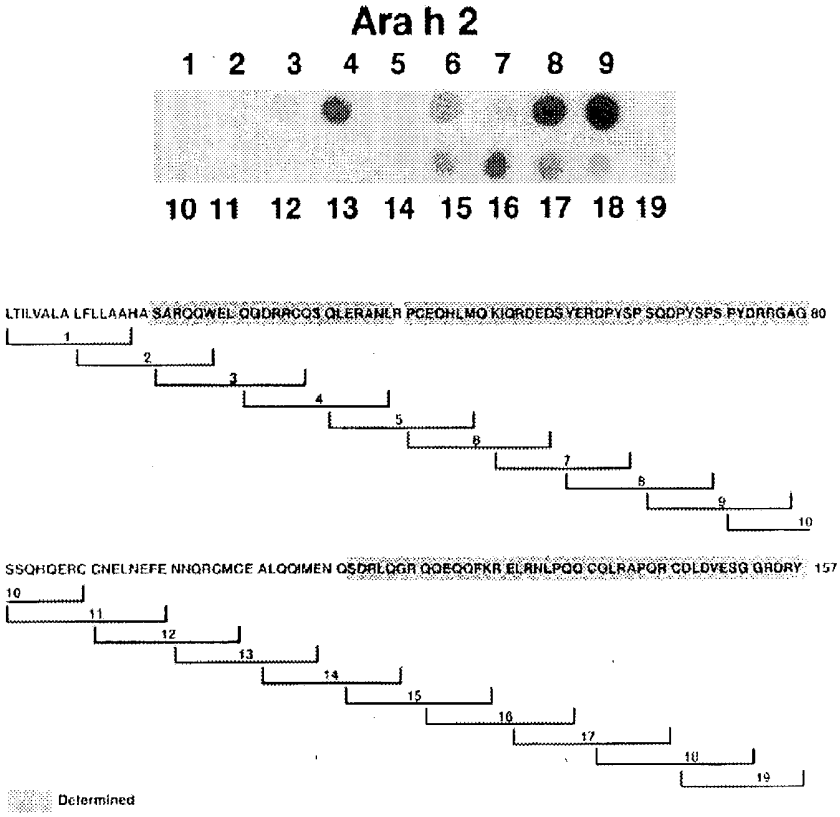


FIG. 2. Multiple IgE-binding sites identified in the *Ara h* 2 allergen. Epitope analysis was performed on the *Ara h* 2 allergen by synthesizing 15 amino acid long peptides, offset from each other by 8 amino acids for the entire protein molecule. These peptides, represented as spots 1–19, were then probed with a serum pool consisting of 15 patients with documented peanut hypersensitivity. The *Ara h* 2 amino acid sequence is shown as the one-letter amino acid code in the bottom half of this figure. The shaded boxes (D1–D3) correspond to the determined IgE binding regions.

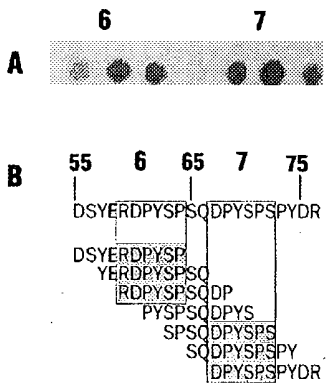


FIG. 3. Core IgE-binding epitopes identified on the *Ara h* 2 allergen. Epitope analysis was performed on the IgE-binding sites identified in Fig. 2 by synthesizing 10-amino-acid-long peptides offset by two amino acids. These peptides were then probed with the 15-patient serum pool. (A) The peptide analysis of *Ara h* 2 amino acid residues 55–76. This region contains peptides 6 and 7 identified in Table III. (B) The amino acid sequence of this region.

seed storage proteins from a variety of different plants (Table II). The highest percent identity (40%) was observed between the *Ara h* 2 protein and conglutin-δ, a sulfur-rich protein from the lupin seed (20). 2S albumins and mabinlins also had a high degree of homology (30–35%) with the *Ara h* 2 protein sequence (21). Interestingly, the *Ara h* 2 protein had some similarity (26–29%) with α-amylase inhibitors from wheat (22, 23), which are the major allergens in baker's asthma (24, 25) and are important allergens in patients experiencing hypersensitivity reactions following the ingestion of wheat protein (26).

Multiple IgE binding epitopes on the Ara h 2 protein. Nineteen overlapping peptides representing the derived amino acid sequence of the *Ara h* 2 protein were synthesized to determine which regions were recognized by serum IgE. Each peptide was 15 amino acids long and was offset from the previous peptide by 8 amino acids. In this manner, the entire length of the *Ara h* 2 protein could be studied in large overlapping fragments. These peptides were then probed with a pool of serum from 15 patients with documented peanut hypersensitivity or serum from a representative control patient with no peanut hypersensitivity. Serum IgE from the control patient did not recognize any of the synthesized peptides (data not shown). In contrast, Fig. 2 shows that there are three IgE binding regions along the entire length of the *Ara h* 2 protein that are recognized by this population of patients with peanut hypersensitivity. These IgE-binding regions represent amino acid residues 17–39, 41–80, and 114–157.

In order to determine the exact amino acid sequence of the IgE binding regions, small peptides (10 amino acids long offset by two amino acids) representing the larger IgE-binding regions were synthesized. In this manner it was possible to identify individual IgE-binding epitopes within the larger IgE-binding regions of the *Ara h* 2 molecule (Fig. 3). The 10 IgE-binding epitopes that were identified in this manner are shown in Table III. The size of the epitopes ranged from 6 to 10 amino acids in length. Three epitopes (aa15-24, aa21-30, and aa27-36), which partially overlapped with each other, were found in the region of amino acid residues 17–39. Four epitopes (aa39-48, aa49-58, aa57-66, and aa65-74) were found in region 41–80. Finally, three epitopes (aa115-124, aa127-136, and aa143-152) were found in the region represented by amino acid residues 114–157. Sixty-three percent of the amino acids represented in the epitopes were either polar uncharged or apolar residues. There was no obvious amino acid sequence motif that was shared by all the epitopes, with the exception of epitopes 6 and 7, which contained the sequence DPYSP.

Identification of the immunodominant Ara h 2 epitopes. In an effort to determine which, if any, of the 10 epitopes was immunodominant, each set of 10 peptides was probed individually with serum IgE from 10 different patients. Five patients were randomly selected from the pool of 15 patients used to identify the common epitopes, and 5 patients were selected from outside this pool. Figure 4A shows an immunoblot strip containing these peptides that has been incubated with an individual patient's serum. This patient's serum IgE recognized peptides 1, 3, 4, 6, and 7. The remaining patients serum IgE were tested in the same manner and the intensity of IgE binding to each spot was determined as a function of that patient's total IgE binding to these 10 epitopes (Fig. 4B) All of the patient sera tested (10/10) recognized multiple peptides. Epitopes 3, 6, and 7 were recognized by serum IgE of all patients tested (10/10). In addition, serum IgE that recognizes these peptides represent the majority of *Ara h* 2-specific IgE found in these patients. These results indicate

TABLE III

Ara h 2 IgE Binding Epitopes

Peptide	AA Sequence	*Ara h* 2 position
1	**HASARQQWEL**	15–24
2	**QWELQGDRRC**	21–30
3	**DRRCQSQLER**	27–36
4	**LRPCEQHLMQ**	39–48
5	**KIQRDEDSYE**	49–58
6	**YERDPYSPSQ**	57–66
7	**SQDPYSPSPY**	65–74
8	**DRLQGRQQEQ**	115–124
9	**KRELRNLPQQ**	127–136
10	**QRCDLDVESG**	143–152

Note. The *Ara h* 2 IgE-binding epitopes (1–10) are indicated as the bold, underlined, single-letter amino acid code. The position of each peptide with respect to the *Ara h* 2 protein is indicated in the right-hand column.

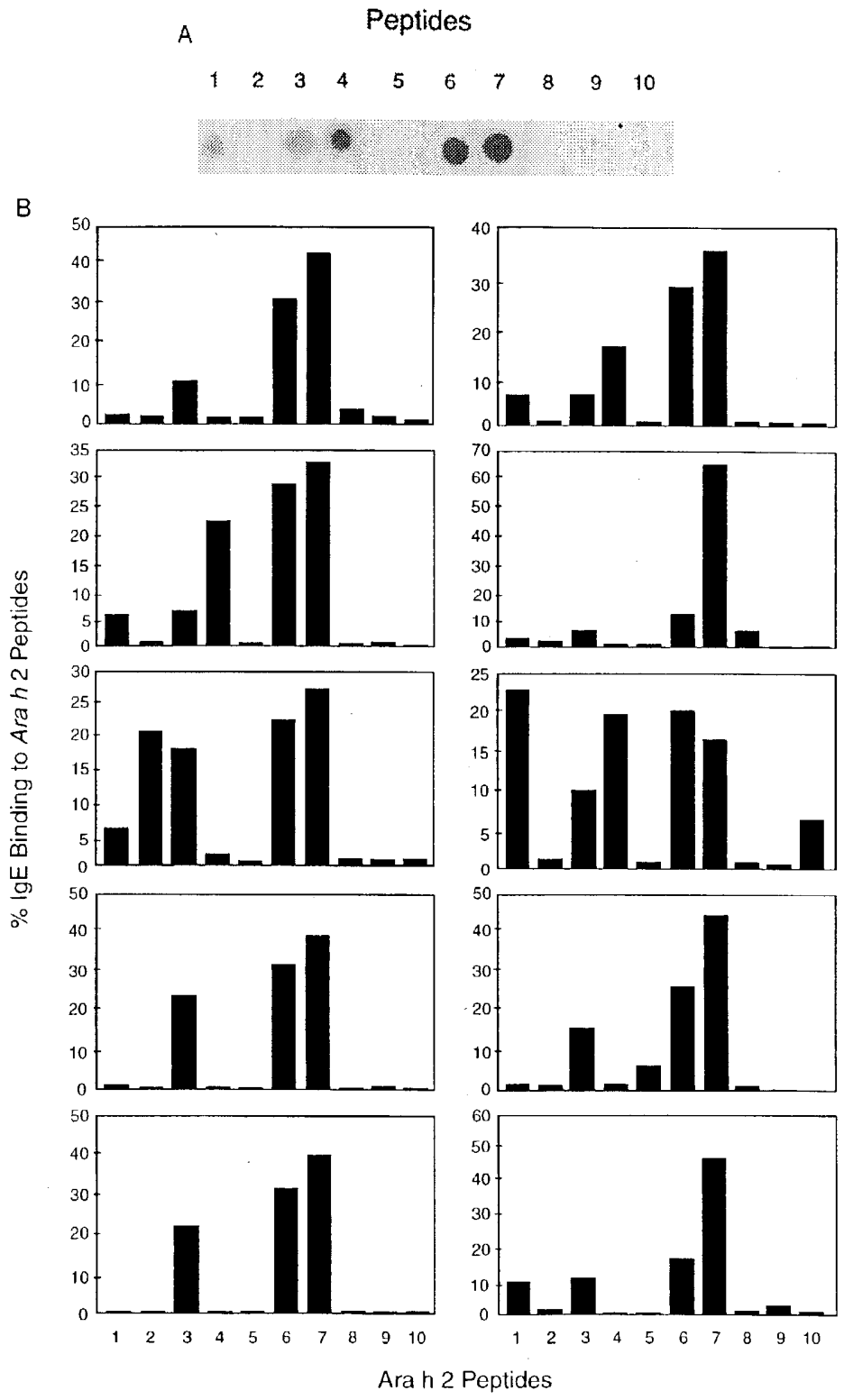


FIG. 4. Identification of the immunodominant *Ara h* 2 epitopes. IgE-binding peptides (1–10) were synthesized as 10 amino acid peptides and probed with serum IgE from 10 patients with documented peanut hypersensitivity. IgE binding to individual peptides was detected by ^{125}I-labeled anti-human IgE. (A) A representative SPOTS membrane containing *Ara h* 2 epitopes 1–10 and probed with serum IgE from a single patient. (B) The relative intensity of IgE binding to each peptide by individual patients was then determined by densitometry and expressed as a percentage of that patient's total IgE binding to all of the *Ara h* 2 peptides. Five patients were selected at random from the 15-patient serum pool and 5 were patients with peanut hypersensitivity that were not included in the pool.

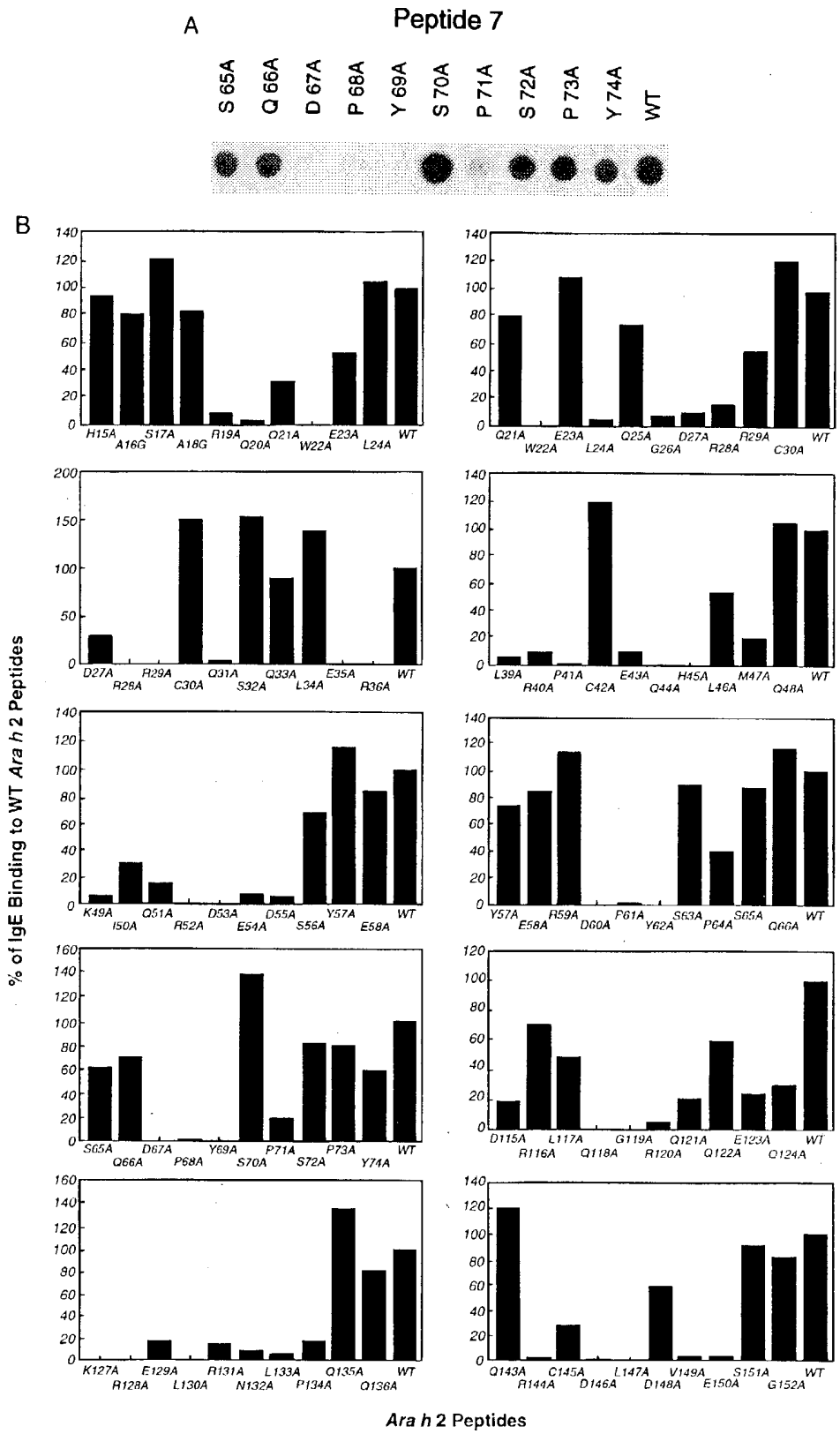


FIG. 5. All of the *Ara h* 2 peptides can be mutated to non-IgE-binding epitopes by single amino acid changes. All of the *Ara h* 2 epitopes (peptides 1–10) were synthesized with an alanine residue substituted for one of the amino acids and then probed with a pool of serum IgE from 15 patients with documented peanut hypersensitivity. IgE binding to individual peptides was detected by ^{125}I-labeled anti-human IgE.

that peptides 3, 6, and 7 contain the immunodominant epitopes of the *Ara h* 2 protein.

The Ara h 2 peptides can be mutated to non-IgE-binding epitopes by single amino acid changes. To assess the importance of individual amino acids in each of the *Ara h* 2 epitopes they were synthesized as 10 amino acid residue peptides with alanine residues being substituted one at a time for each of the amino acids in the epitope. These peptides were then probed with pooled serum IgE from 15 patients with documented peanut hypersensitivity. Figure 5A shows an immunoblot strip containing the wild-type and mutated peptides of epitope 7. The pooled serum IgE did not recognize this peptide or binding was drastically reduced when alanine was substituted for amino acids at position 67, 68, or 69. In contrast, the substitution of an alanine for a serine residue at position 70 resulted in increased IgE binding. The remaining *Ara h* 2 epitopes were tested in the same manner and the intensity of IgE binding to each spot was determined as a percentage of IgE binding to the wild-type peptide (Fig. 5B). In general, each epitope could be mutated to a non-IgE-binding peptide by the substitution of an alanine for a single amino acid residue. There was no obvious position within each peptide that, when mutated, would result in loss of IgE binding. Furthermore, there was no consensus in the type of amino acid that, when changed to alanine, would lead to loss of IgE binding.

DISCUSSION

Peanuts are one of the most common food allergens in both children and adults. In addition, peanut hypersensitivity is less likely to resolve spontaneously and more likely to result in fatal anaphylaxis. Because of the significance of the allergic reaction and the widening use of peanuts as protein extenders in processed food, the risk to the peanut-sensitive individual is increasing.

Various studies over the last several years have examined the nature and location of the multiple allergens in peanuts (8–10). Taylor *et al.* (9) demonstrated that the allergenic portion of peanuts was in the protein portion of the cotyledon. Our laboratory recently identified two major allergens from peanut extracts, designated *Ara h* 1 (11) and *Ara h* 2 (12). Greater than 90% of our patients who were challenge positive to peanut had specific IgE to these proteins. The *Ara h* 1 allergen has been identified as a seed storage protein with sequence homology to the vicilins (13), a family of proteins commonly found in many higher plants (27, 28). The *Ara h* 2 nucleotide sequence identified in this report has sequence homology with another class of seed storage proteins called conglutins (20). It is interesting to note that two of the major peanut allergens thus far identified are seed storage proteins that have sequence homology with proteins in other plants. This may explain the cross-reacting antibodies to other legumes that are found in the sera of patients that manifest clinical symptoms to only one member of the legume family (29).

There are at least 10 IgE recognition sites distributed throughout the major peanut allergen *Ara h* 2. The identification of multiple epitopes on a single allergen is not novel, there being reports of multiple IgE binding epitopes on allergens from many foods that cause immediate hypersensitivity reactions (30–36). The observation that most of these proteins have multiple IgE-binding sites probably reflects the polyclonal nature of the immune response to them and may be a necessary step in establishing a protein as an allergen.

Recent evidence suggests that there is a preferential variable heavy chain usage in IgE synthesis and a direct switching from IgM production to IgE synthesis (37). This would suggest that epitopes responsible for antigen-specific IgE antibody production may differ from those promoting antigen-specific IgG antibodies and that there may be some structural similarity between peptides that elicit IgE antibody production. However, there was no obvious sequence motif that was shared by the 23 different IgE-binding epitopes of the peanut allergen *Ara h* 1 (13). In the present study, two epitopes in *Ara h* 2 share a hexameric peptide (DPYSPS). It is significant to note that these peptides are recognized by serum IgE from all the peanut hypersensitive patients tested in this study. In addition, serum IgE that recognize these peptides represent the majority of *Ara h* 2-specific IgE found in these patients. Whether there is any further structural similarity between the IgE binding epitopes of *Ara h* 2 remains to be determined.

The elucidation of the major IgE binding epitopes on *Ara h* 2 may enable us to design better therapeutic options for the prevention of anaphylaxis as a result of peanut hypersensitivity. The only therapeutic option presently available for the prevention of a food hypersenitivty reaction is food avoidance. Unfortunately, for a ubiquitous food such as peanut, the possibility of an

(A) A representative SPOTS membrane containing the wild-type (WT) peptide 7 and the mutated forms in which the amino acid at each position was substituted with alanine. (B) The relative intensity of IgE binding to the mutated forms of each peptide was then determined by densitometry and expressed as a percentage of IgE binding to the wild-type peptide. The letters and numbers across the top of (A) and across the bottom of each panel in (B) indicate the one-letter amino acid code for the residue normally at that position, the position of each residue in the *Ara h* 2 protein, and the amino acid that was substituted at that position. WT, indicates the wild-type peptide (no amino acid substitutions).

inadvertent ingestion is great. One therapeutic option used extensively for patients with allergic reactions to various aeroallergens and insect sting venoms is allergen desensitization immunotherapy. Allergen immunotherapy consists of injections of increasing amounts of allergens to which a patient has Type I immediate hypersensitivity (38, 39). While the absolute mechanism of immunotherapy is unknown, an increase in IgG or IgG_4 antibody activity, a decrease in allergen-specific IgE levels, and a decrease in basophil activity have all been implicated in mediating this response (40–43). Because allergen immunotherapy has been proven efficacious for treatment of some allergies, treatment with peanut immunotherapy is now being studied as a possible option (44). Our work showing that the IgE-binding epitopes of a major peanut allergen may allow for the use of immunodominant epitopes in this approach.

Finally, our data show that it may be possible to mutate the *Ara h* 2 allergen to a protein that no longer binds IgE. This raises the possibility that an altered *Ara h* 2 gene could be used to replace its allergenic homologue in the peanut genome. In this manner, a hypoallergenic peanut could be developed that may blunt allergic reactions in sensitive individuals who inadvertently ingest this food. Such an approach would require that the altered *Ara h* 2 gene product retain its native function and properties. Given the potential severity of peanut allergic reactions and the widespread use of peanuts in consumer foods, this possibility is actively being pursued in our laboratories.

ACKNOWLEDGMENTS

This work was supported by grants from the Arkansas Science and Technology Authority, National Institute of Health Grant RO1-AI33596, and the Clarissa Sosin Allergy Research Foundation.

REFERENCES

1. Burks, A. W., and Sampson, H. A. (1993) *Curr. Prob. Pediatr.* **23,** 230–252.
2. Jansen, J. J. N., Kardinaai, A. F. M., Huijbers, G. H., Vlieg-Boerstra, B. J., Martens, B. P. M., and Ochuizen, T. (1994) *J. Allergy Clin. Immunol.* **93,** 446–456.
3. Yunginger, J. W., and Jones, R. T. (1987) A review of peanut chemistry: Implications for the standardization of peanut extracts. *in* Regulatory Control and Standardization of Allergenic Extracts: Proceedings of the Fourth International Paul Ehrlich Seminar. (Schaeffer, M., Sisk, C., and Brede, H. D., Eds.), pp. 251–264, Washington, DC.
4. Bock, S. A. (1985) *J. Pediatr.* **107,** 676–680.
5. Sampson, H. A., Mendelson, I., and Rosen, J. P. (1992) *N. Engl. J. Med.* **327,** 380–384.
6. Yunginger, J. W., Sweeney, K. G., Sweeney, W. Q., Giannandrea, L. A., Teigland, J. D., Bray, M., Benson, P. A., York, J. A., Biedrzycki, L., Squillace, D. L., and Helm, R. M. (1988) *J. Am. Med. Assoc.* **260,** 1450–1452.
7. Yunginger, J. W., Squillace, D. L., Jones, R. T., and Helm, R. M. (1989) *Allergy Proc.* **10,** 249–253.
8. Bush, R. K., Taylor, S. L., and Nordlee, J. A. (1989) *Allergy Proc.* **10,** 261–264.
9. Taylor, S. L., Busse, W. W., Sachs, M. I., Parker, J. L., and Yunginger, J. W. (1981) *J. Allergy Clin. Immunol.* **68,** 373–375.
10. Barnett, D., Baldo, B. A., and Howden, E. H. (1983) *J. Allergy Clin. Immunol.* **72,** 61–68.
11. Burks, A. W., Williams, L. W., Helm, R. M., Connaughton, C., Cockrell, G., and O'Brien, T. J. (1991) *J. Allergy Clin. Immunol.* **88,** 172–179.
12. Burks, A. W., Williams, L. W., Connaughton, C., Cockrell, G., O'Brien, T. J., and Helm, R. M. (1992) *J. Allergy Clin. Immunol.* **90,** 962–969.
13. Burks, A. W., Cockrell, G., Stanley, J. S., Helm, R. M., and Bannon, G. A. (1996) *J. Clin. Invest.* **96,** 1715–1721.
14. Larsen, J. N., Stroman, P., and Ipsen, H. (1992) *Mol. Immunol.* **29,** 703–711.
15. Bannon, G. A., Calzone, F. J., Bowen, J. K., Allis, C. D., and Gorovsky, M. A. (1983) *Nucleic Acids Res.* **11,** 3903–3917.
16. Needleman, S. B., and Wunsch, C. D. (1970) *J. Mol. Biol.* **48**(3), 443–453.
17. Watson, C. J., and Jackson, R. F. (1985) DNA Cloning (D. M. Glover, Ed.), Vol. I, pp. 79–88, IRL Press.
18. Huynh, T. Y., Young, R. A., and Davies, R. W. (1985) *in* DNA Cloning, A Practical Approach (Glover, D. M., Ed.), Vol. I, pp. 49–78, Oxford Press.
19. Sanger, F., Nicklen, S., and Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
20. Gayler, K. R., Kolivas, S., Macfarlane, A. J., Lilley, G. G., Baldi, M., Blagrove, R. J., and Johnson, E. D. (1990) *Plant Mol. Biol.* **15,** 879–893.
21. Nirasawa, S., Nishino, T., Katahira, M., Uesugi, S., Hu, Z., and Kurihara, Y. (1994) *Eur. J. Biochem.* **223,** 989–995.
22. Garcia-Maroto, F., Marana, C., Mena, M., Garcia-Olmedo, F., and Carbonero, P. (1990) *Plant Mol. Biol.* **14,** 845–853.
23. Joudrier, P., Lullien-Pellerin, V., Alary, R., Grosset, J., Guirao, A., and Gautier, M. F. (1995) *DNA Seq.* **5,** 153–162.
24. Sanchez-Monge, R., Gomez, L., Barber, D., Lopez-Otin, C., Armentia, A., and Salcedo, G. (1992) *Biochem. J.* **281,** 401–405.
25. Armentia, A., Sanchez-Monge, R., Gomez, L., Barber, D., and Salcedo, G. (1993) *Clin. Exp. Allergy* **23,** 410–415.
26. James, J. M., Sixbey, J. P., Helm, R. M., Bannon, G. A., and Burks, A. W. (1996) *J. Allergy Clin. Immunol.,* in press.
27. Gibbs, P. E., Strongin, K. B., and McPherson, A. (1989) *Mol. Biol. Evol.* **6,** 614–623.
28. Chee, P. P., and Slightom, J. L. (1991) *Sub-Cell. Biochem.* **17,** 31–52.
29. Bernhisel-Broadbent, J., Taylor, S., and Sampson, H. A. (1989) *J. Allergy Clin. Immunol.* **84,** 701–709.
30. van Kampen, V., Becker, W. M., Chen, Z., Rihs, H. P., Mazur, G., Raulf, M., Liebers, V., Isringhoausen-Bley, S., and Baur, X. (1994) *Mol. Immunol.* **31,** 1133–1140.
31. Breiteneder, H., Ferreira, F., Reikerstorfer, A., Duchene, M., Valenta, R., Hoffman-Sommergruber, K., Ebner, C., Breitenbach, M., Kraft, D., and Scheiner, O. (1992) *J. Allergy Clin. Immunol.* **90,** 909–917.
32. Ball, G., Shelton, M. J., Walsh, B. J., Hill, D. J., Hoskins, C. S., and Howden, M. E. (1994) *Clin. Exp. Allergy* **24,** 758–764.
33. Aas, K., and Elsayed, S. (1975) *Dev. Biol. Stand.* **29,** 90–98.
34. Elsayed, S., Holen, E., and Dybenal, T. (1989) *Int. Arch. Allergy Appl. Immunol.* **89,** 410–415.
35. Herian, A. M., Taylor, S. L., and Bush, R. K. (1990) *Int. Arch. Allergy Appl. Immunol.* **92,** 193–198.

36. Shanti, K. N., Martin, B. M., Nagpal, S., Metcalf, D. D., and Rao, P. V. *J. Immunol.* **151,** 5354–5363.
37. van der Stoep, N., Korvere, W., and Logtenberg, T. (1994) *Eur. J. Immunol.* **24,** 1307–1311.
38. Reisman, R. E. (1994) *Allergy Proc.* **15,** 61–63.
39. Fitzsimons, T., and Grammer, L. C. (1990) *Allergy Proc.* **11,** 156.
40. Sparholt, S. H., Olsen, O. T., and Schou, C. (1992) *Clin. Exp. Allergy* **22,** 648–653.
41. Gieni, R. S., Yang, X., and Hayglass, K. T. (1993) *J. Immunol.* **150,** 302–310.
42. Secrist, H., Chelen, C. J., Wen, Y., Marshall, J. D., and Umetsu, D. T. (1993) *J. Exp. Med.* **178,** 2123–2130.
43. Garcia, N. M., Lynch, M. C., Di Prisco, M. C., and Lopez, R. I. (1995) *J. Invest. Allergol. Clin. Immunol.* **5,** 18–24.
44. Oppenheimer, J. J., Nelson, H. S., Bock, S. A., Christensen, F., and Leung, D. Y. (1992) *J. Allergy Clin. Immunol.* **90,** 151–152.

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

From genes to protein structure and function: novel applications of computational approaches in the genomic era

Jeffrey Skolnick and Jacquelyn S. Fetrow

The genome-sequencing projects are providing a detailed 'parts list' of life. A key to comprehending this list is understanding the function of each gene and each protein at various levels. Sequence-based methods for function prediction are inadequate because of the multifunctional nature of proteins. However, just knowing the structure of the protein is also insufficient for prediction of multiple functional sites. Structural descriptors for protein functional sites are crucial for unlocking the secrets in both the sequence and structural-genomics projects.

Genome-sequencing projects are providing a detailed 'parts list' for life. Unfortunately, this list, a portion of which represents the amino acid sequence of all the proteins in a given genome, does not come with an instruction manual. That is, given the genome's sequences, one does not necessarily know straight away which regions encode proteins, which serve a regulatory role and which are responsible for the structure and replication of the DNA itself.

This is not unlike giving a child a list of parts necessary to create a working automobile. Without the necessary expertise, creating the final, working car from just the initial parts list is a nearly impossible task. Similarly, understanding how to create a complete, functioning cell given just the sequence of nucleotides found in an organism's genome is a complex problem.

What is a protein function?

After a genome is sequenced and its complete parts list determined, the next goal is to understand the function(s) of each part, including that of the proteins. What do we mean by protein function, the focus of this article?

Function has many meanings. At one level, the protein could be a globular protein, such as an enzyme, hormone or antibody, or it could be a structural or membrane-bound protein. Another level is its biochemical function, such as the chemical reaction and the substrate specificity of an enzyme. The regulatory molecules or cofactors that bind to a protein are also levels of biochemical function.

At the cellular level, the protein's function would involve its interaction with other macromolecules and the function and cellular location of such complexes. There is also the protein's physiological function; that is, in which metabolic pathway the protein is involved or what physiological role it performs in the organism. Finally, the phenotypic function is the role played by the protein in the total organism, which is observed by deleting or mutating the gene encoding the protein.

J. Skolnick (skolnick@danforthcenter.org) is at the Danforth Plant Science Center, Laboratory of Computational Genomics, 4041 Forest Park Avenue, St Louis, MO 63108, USA. J.S. Fetrow is at GeneFormatics, Suite 200, 5830 Oberlin Drive, San Diego, CA 92121-3754, USA.

Obviously, the complete characterization of protein function is difficult but efforts are under way at all levels[1–4], including cellular function[5,6]. In this article, however, we focus on identifying the biochemical function of a protein given its sequence, a problem that is amenable to molecular approaches.

Sequence-based approaches to function prediction

The sequence-to-function approach is the most commonly used function-prediction method. This robust field is well developed and, in the interest of space limitations, we will merely present a brief overview.

There are two main flavors of this approach: sequence alignment[7–9]; and sequence-motif methods such as Prosite[10], Blocks[11], Prints[12,13] and Emotif[14]. Both the alignment and the motif methods are powerful but a recent analysis has demonstrated their significant limitations[15], suggesting that these methods will increasingly fail as the protein-sequence databases become more diverse.

An extension of these approaches that combines protein-sequence with structural information has been developed and some successes have been reported[16]. However, this method still applies the structural information in a one-dimensional, 'sequence-like' fashion and fails to take into account the powerful three-dimensional information displayed by protein structures.

In addition, proteins can gain and lose function during evolution and may, indeed, have multiple functions in the cell (Box 1). Sequence-to-function methods cannot specifically identify these complexities. Inaccurate use of sequence-to-function methods has led to significant function-annotation errors in the sequence databases[17].

An alternative approach

An alternative, complementary approach to protein-function prediction uses the sequence-to-structure-to-function paradigm. Here, the goal is to determine the structure of the protein of interest and then to identify the functionally important residues in that structure. Using the chemical structure itself to identify functional sites is more in line with how the protein actually works.

0167-7799/00/$ – see front matter PII: S0167-7799(99)01398-0

In a sense, this is one long-term goal of 'structural genomics' projects[18,19], which are designed to determine all possible protein folds experimentally, just as genome-sequencing projects are determining all protein sequences[20]. This is in contrast to traditional structural-biology approaches, in which one knows the protein's function first and only then, if the function is sufficiently important, determines its structure.

It is implicitly assumed that having the protein's structure will provide insights into its function, thereby furthering the goals of the human-genome-sequencing project. However, knowing a protein's three-dimensional structure is insufficient to determine its function (Box 2). What we really need to analyse and predict the multifunctional aspects of proteins is a method specifically to recognize active sites and binding regions in these protein structures.

Active-site identification

In order to use a structure-based approach to function prediction, one must identify the key residues responsible for a given biochemical activity. For many years, it has been suggested that the active sites in proteins are better conserved than the overall fold. Taken to the limit, this suggests that one could not only identify distant ancestors with the same global fold and the same activity but also proteins with similar functions but distantly related, or possibly unrelated, global folds.

The validity of this suggestion was demonstrated empirically by Nussinov and co-workers, who showed that the active sites of eukaryotic serine proteases, subtilisins and sulfhydryl proteases exhibit similar structural motifs[21]. Furthermore, in a recent modeling study of *Saccharomyces cerevisiae* proteins, protein functional sites were found to be more conserved than other parts of the protein models[22]. Similarly, it has been demonstrated that the catalytic triad of the α/β hydrolases is structurally better conserved than other histidine-containing triads[23]. A comparison of the structure of the hydrolase catalytic triad to other histidine-containing triads shows a distinct bimodal distribution, while a similar analysis done with a randomly selected triad shows a unimodal distribution (Fig. 1).

Kasuya and Thornton[24] generalized this example by creating structural analogs of a few Prosite sequence motifs[10]. For the 20 most-frequently occurring Prosite patterns, the associated local structure is quite distinct. These results provide clear evidence that enzyme active sites are indeed more highly conserved than other parts of the protein.

Identifying active sites in experimental structures

Historically, several groups have attempted to identify functional sites in proteins; these efforts were directed at protein engineering or building functional sites in places where they did not previously exist. This has been successfully accomplished for several metal-binding sites[25–33]. However, highly accurate functional-site descriptors of the backbone and side-chain atoms were required, fueling the belief that significant atomic detail is required in site descriptors for function identification.

Highly detailed residue side-chain descriptors of the active sites of serine proteases and related proteins have been used to identify functional sites[3]. The use of these highly detailed motifs has led to the identification of several novel functional sites in known, high-quality protein structures[3,34]. More automated methods for finding spatial motifs in protein structures have also been described[21,34–40].

Unfortunately, most of these methods require the exact placement of atoms within protein backbones and side chains, and so have not been shown to be relevant to inexact predicted structures. Recently, however, we described the production of fuzzy, inexact descriptors of protein functional sites[15]. As we wish to apply the descriptors to experimental structures as well as to predicted protein models, we used only carbon atoms and side-chain centers-of-mass positions. We call these descriptors 'fuzzy functional forms' (FFFs) and have created them for both the disulfide-oxidoreductase[15,41] and α/β-hydrolase catalytic active sites[23].

The disulfide-oxidoreductase FFF was applied to screen high-resolution structures from the Brookhaven protein database[42]. In a dataset of 364 protein structures, the FFF accurately identified all proteins known to exhibit the disulfide-oxidoreductase active site[15]. In a larger dataset of 1501 proteins, the FFF again accurately identified all proteins with the active site. In addition, it identified another protein, 1fjm, a serine–threonine phosphatase. This result was initially discouraging but subsequent sequence alignment and clustering analysis strongly suggested that this putative site might indeed be a site of redox regulation in the serine–threonine phosphatase-1 subfamily[43]. If confirmed by experiment, this result will highlight the advantages of using structural descriptors to analyse multiple functional sites in proteins. It will also highlight the fact that human

Box 1. Proteins are multifunctional

A common protein characteristic that makes functional analysis based only on homology especially difficult is the tendency of proteins to be multifunctional. For instance, lactate dehydrogenase binds NAD, substrate and zinc, and performs a redox reaction. Each of these occurs at different functional sites that are in close proximity and the combination of all four sites creates the fully functional protein.

Other examples of multifunctional proteins are the nucleic-acid-binding proteins. For instance, DNA regulatory proteins often contain a DNA-binding domain, a multimerization domain and additional sites that bind regulatory proteins; a classic example is RecA[59]. The 3C rhinovirus protease exhibits a proteolytic function as well as an RNA-binding function[60,61]. Transcription factors are also complex, multifunctional proteins[62]. It is becoming increasingly important to recognize each of these different functions of gene products of a newly sequenced gene.

The serine–threonine-phosphatase superfamily is a prime example of the difficulties of using standard sequence analysis to recognize the multiple functions found in single proteins. This large protein family is divided into a number of subfamilies, all of which contain an essential phosphatase active site. Subfamilies 1, 2A and 2B exhibit 40% or more sequence identity between them[63]. However, each of these subfamilies is apparently regulated differently in the cell[64–67] and observation suggests that there are different functional sites at which regulation can occur. Because the sequence identity between subfamilies is so high, standard sequence-similarity methods could easily misclassify new sequences as members of the wrong subfamily if the functional sites are not carefully considered, as was recently demonstrated[43].

These are but a few examples of the multifunctionality of proteins. The recognition of this multifunctional nature is of critical importance to the genomics field. Useful functional-annotation methods must consider all of the specific functions in a given protein and will not just provide a general classification of function.

Box 2. Knowing a protein's structure does not necessarily tell you its function

Because proteins can have similar folds but different functions[68,69], determining the structure of a protein may or may not tell you something about its function. The most well-studied example is the $(\alpha/\beta)_8$ barrel enzymes, of which triose-phosphate isomerase (TIM) is the archetypal representative. Members of this family have similar overall structures but different functions, including different active sites, substrate specificities and cofactor requirements[70,71].

Is this example common? Our own analysis of the 1997 SCOP database[68] shows that the five largest fold families are the ferredoxin-like, the (α/β) barrels, the knottins, the immunoglobulin-like and the flavodoxin-like fold families with 22, 18, 13, 9 and 9 subfamilies, respectively (Fig. i). In fact, 57 of the SCOP fold families consist of multiple superfamilies. These data only show the tip of the iceberg, because each superfamily is further composed of protein families and each individual family can have radically different functions. For example, the ferredoxin-like superfamily contains families identified as Fe–S ferredoxins, ribosomal proteins, DNA-binding proteins and phosphatases, among others.

After this article was submitted, a much-more-detailed analysis of the SCOP database was published[72]. This finds a broad function–structure correlation for some structural classes, but also finds a number of ubiquitous functions and structures that occur across a number of families. The article provides a useful analysis of the confidence with which structure and function can be correlated[72]. Knowing the protein structure by itself is insufficient to annotate a number of functional classes and is also insufficient for annotating the specific details of protein function.

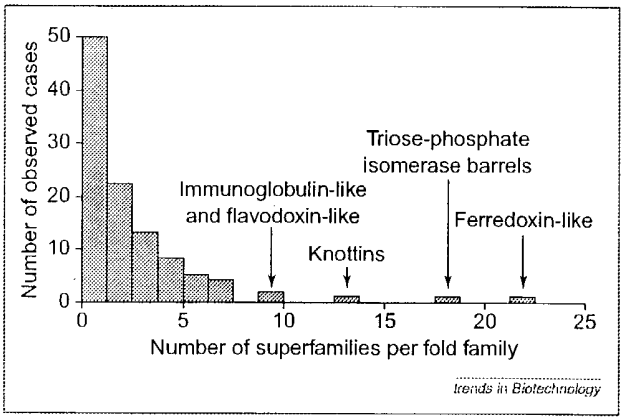


Figure i

Histogram of the numbers of superfamilies found in each SCOP fold family. These data clearly show that proteins with similar structures can have different functions and demonstrate the difficulty of assigning protein function based simply on the three-dimensional structure. The data were taken from the 1997 distribution of SCOP (http://scop.mrc-lmb.cam.ac.uk/scop). For a more-detailed analysis, see Ref. 72.

observation alone is no longer adequate for identifying all functional sites in known protein structures.

To date, the use of structure to identify function has largely focused on high-resolution structures and highly detailed descriptors of protein functional sites. However, the creation of inexact descriptors for functional sites opens the way to the application of these methods to inexact, predicted protein models. The question remains: how good does a model have to be in order to use FFFs to identify its active sites?

The state of the art in structure-prediction methods

For proteins whose sequence identity is above ~30%, one can use homology modeling to build the structure[44]. However, structure prediction is far more difficult for proteins that are not homologous to proteins with known structure. At present, there are two approaches for these sequences: *ab initio* folding[45–48] and threading[49–53].

In *ab initio* folding, one starts from a random conformation and then attempts to assemble the native structure. As this method does not rely on a library of pre-existing folds, it can be used to predict novel folds. The recent CASP3 protein-structure-prediction experiment (http://PredictionCenter.llnl.gov/CASP3) involved the blind prediction of the structure of proteins whose actual structure was about to be experimentally determined. These results indicate that considerable progress has been made[46,54]. For helical and α/β proteins with less than 110 residues, structures were often predicted whose backbone root-mean-square deviation (RMSD) from native ranged from 4–7 Å. Progress is being made with the β proteins, too, although they remain problematic. Because *ab initio* methods can identify novel folds, these methods could be used to help to select sequences likely to yield novel folds in experimental structural-genomics projects.

Another approach to tertiary-structure prediction is threading. Here, for the sequence of interest, one attempts to find the closest matching structure in a library of known folds[52,55]. Threading is applicable to proteins of up to 500 residues or so and is much faster than *ab initio* approaches. However, threading cannot be used to obtain novel folds.

Ab initio *predicted models can be used for automatic protein-function prediction*

The results of the recent CASP3 competition suggest that current modeling methods can often (but not always) create inexact protein models. Are these structures useful for identifying functional sites in proteins? Using the *ab initio* structure-prediction program MONSSTER, the tertiary structure of a glutaredoxin, 1ego, was predicted[56]. For the lowest-energy model, the overall backbone RMSD from the crystal structure was 5.7 Å.

To determine whether this inexact model could be used for function identification, the sets of correctly and incorrectly folded structures were screened with the FFF for disulfide-oxidoreductase activity[15]. The FFF uniquely identified the active site in the correctly folded structure but not in the incorrectly folded ones (Fig. 2). This is a proof-of-principle demonstration that inexact models produced by *ab initio* prediction of structure from sequence can be used for the subsequent prediction of biochemical function. Of course, improvements in the method have to be made before such predictions can be done on a routine basis.

Use of predicted structures from threading in protein-function prediction

At present, practical limitations preclude folding an entire genome of proteins using *ab initio* methods[57]. Threading is more appropriate for achieving the requisite high-throughput structure prediction. Thus, a standard threading algorithm[58] has been used to screen all

proteins in nine genomes for the disulfide-oxidoreductase active site described above.

First, sequences that aligned with the structures of known disulfide oxidoreductases were identified. Then, the structure was searched for matches to the active-site residues and geometry. For those sequences for which other homologs were available, a sequence-conservation profile was constructed[23]. If the putative active-site residues were not conserved in the sequence subfamily to which the protein belongs, that sequence was eliminated. Otherwise, the sequence is predicted to have the function.

Using this sequence-to-structure-to-function method, 99% of the proteins in the nine genomes that have known disulfide-oxidoreductase activity have been found. From 10% to 30% more functional predictions are made than by alternative sequence-based approaches; similar results are seen for the α/β hydrolases[23]. Surprisingly, in spite of the fact that threading algorithms have problems generating good sequence-to-structure alignments, active sites are often accurately aligned, even for very distant matches. This observation would agree with the above experimental results indicating that active sites are well conserved in protein structures.

Importantly, the false-positive rate when using structural information is much lower than that found using sequence-based approaches, as demonstrated by a detailed comparison of the FFF structural approach and the Blocks sequence-motif approach (N. Siew *et al.*, unpublished). In this study, the sequences in eight genomes, including *Bacillus subtilis*, were analysed for disulfide-oxidoreductase function using the disulfide-oxidoreductase FFF, the thioredoxin Block 00194 and the glutaredoxin Block 00195. If we assume that those sequences identified by both the FFF and Blocks are 'true positives', we find 13 such sequences in the *B. subtilis* genome.

There is no experimental evidence validating all of these 'true positives' and so they are more accurately termed 'consensus positives'. In order to find these 13 'consensus positive' sequences, the FFF hits seven false positives. On the other hand, Blocks hits 23 false positives (Fig. 3). It was previously suggested that the use of a functional requirement adds information to threading and reduces the number of false positives[52]. These data, including the data shown in Fig. 3, validate this claim on a genome-wide basis.

Of course, as no genome has had the function of all of its proteins experimentally annotated, it is impossible to know how many other proteins with the specified biochemical function were not properly identified. This is a critical question for researchers attempting to predict protein function. Experimental confirmation will be needed to validate this or any other method fully. This points out the need for closely coupling computational function-prediction algorithms with experiments.

Weaknesses of using the sequence-to-structure-to-function method of function prediction

Based on studies to date, the identification of enzymatic activity requires a model in which the backbone RMSD from native near the active sites is about 4–5 Å. Predicted models are better at describing the geometry in the core of the molecule than in the loops and so predicting the function of a protein whose active site is in loops may be a problem. Also, the method can currently only be applied to enzyme active sites; substrate- and ligand-binding sites have not been identified using the inexact models. Techniques that will further refine inexact protein models will be quite useful in taking the protein analysis to the next step.

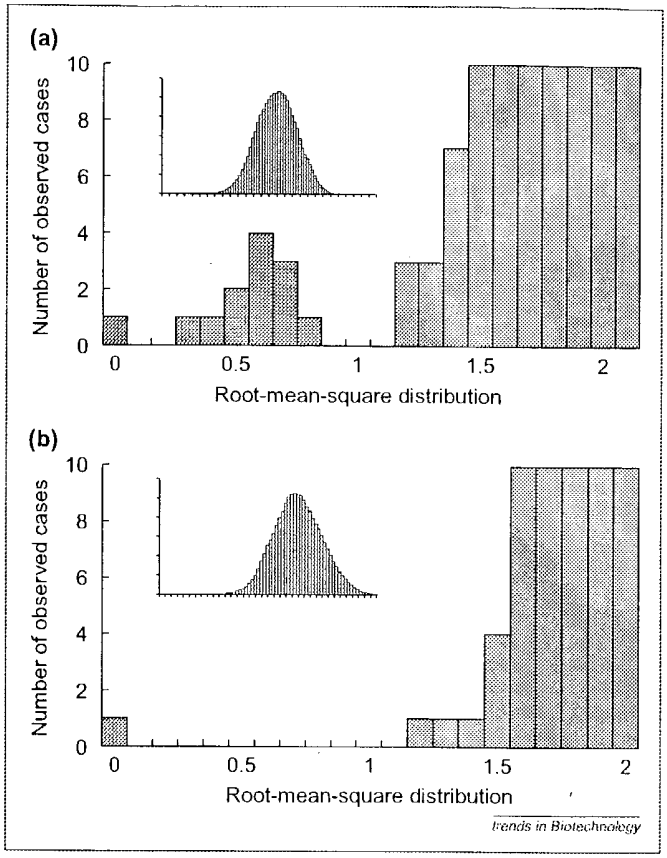


Figure 1

The distribution of root-mean-square distributions (RMSD) between the hydrolase catalytic triad and all other histidine-containing triads shows a bimodal distribution **(a)**; by contrast, the RMSD between a randomly selected (non-catalytic) triad and all other histidine-containing triads has a unimodal distribution **(b)**. The His–Ser–Asp catalytic trial in the protein-1 gpl (Rp2 lipase) **(a)** and a random histidine-containing triad from 4pga (glutaminase–asparaginase) **(b)** were structurally aligned to all His-containing triads in a database of 1037 proteins[23]. Actual α/β-hydrolase active sites **(a)** and the 4pga site **(b)** are indicated by blue bars; other histidine triads that are not active sites are indicated by red bars. None of the sites found by matching to the 4pga were hydrolase active sites. Inset graphs show the full distribution.

Conclusions

Although sequence-based approaches to protein-function prediction have proved to be very useful, alternatives are needed to assign the biochemical function of the 30–50% of proteins whose function cannot be assigned by any current methods. One emerging approach involves the sequence-to-structure-to-function paradigm. Such structures might be provided by structural-genomics projects or by structure-prediction algorithms. Functional assignment is made by screening the resulting structure against a library of structural descriptors for known active sites or binding regions.

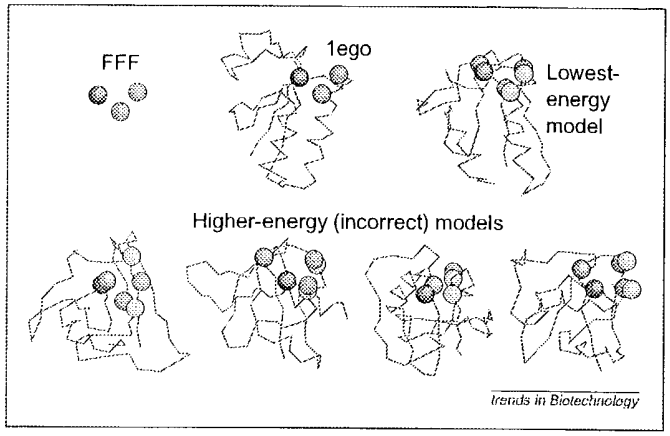


Figure 2

Application of the disulfide-oxidoreductase fuzzy functional form (FFF) to *ab initio* models of glutaredoxin created by the program MONSSTER shows that the FFF can distinguish between correctly folded and misfolded (or higher-energy) models. The FFF is shown as two orange balls (representing the cysteines) and a blue ball (representing the proline). The protein models are shown as magenta wire models with the active-site cysteines and proline shown as yellow and cyan balls, respectively. The FFF clearly distinguishes the correct active site in the crystal structure of the glutaredoxin 1ego and the correctly folded, lowest-energy model. The FFF does not match to the active sites of any of the higher energy, misfolded structures, four of which are shown here.

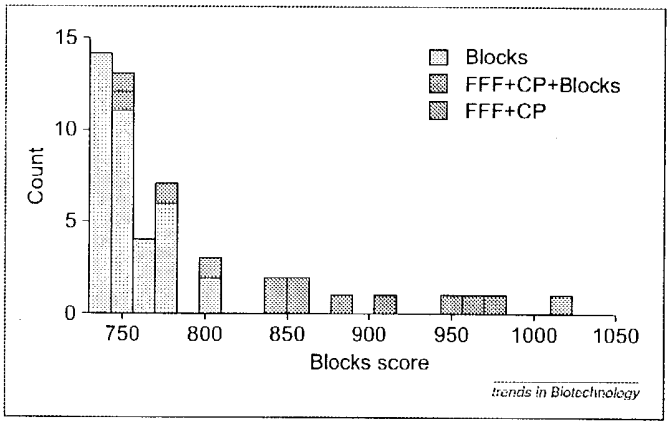


Figure 3

Analysis of the *Bacillus subtilis* genome using the thioredoxin Block 00194. The Blocks score (computed using the publicly available BLIMPS program) is plotted on the *x* axis and the number of sequences found in each scoring bin is plotted on the *y* axis. Those sequences identified as 'consensus positives' [identified by both the fuzzy functional form (FFF) and the Block] are shown as red bars. One additional sequence found by the FFF, which is likely to be a true positive, is shown as a blue bar. All other sequences, putative 'false positives', are shown as yellow bars. Using the Blocks score at which all 13 of the 'consensus positives' are found, 23 false positives are also found. In its analysis of the *B. subtilis* genome, the FFF identifies only seven false positives along with the same 13 'consensus positives' (data not shown).

Detailed descriptors will only work on the experimentally determined, high-quality structures. Ideally, however, the descriptors should work on both experimental structures and the cruder models provided by tertiary-structure-prediction algorithms.

The advantages of such an approach are that one need not establish an evolutionary relationship in order to assign function, that more than one function can be assigned to a given protein [an issue of major importance, because proteins are multifunctional (Box 1)] and, ultimately, that having a structure can provide deeper insight into the biological mechanism of protein function and regulation. The disadvantages are that one needs to have the protein's structure before a function can be assigned and that the approach is limited to those functions associated with proteins with at least one solved structure, so that a functional-site descriptor can be constructed.

In this sense, structure-to-function assignment can be thought of as 'functional threading' – find the active-site match in a library of descriptors for known protein active sites. This is the first step in the long process of using structure to assign all levels of function, a goal that is made increasingly important with the emergence of structural genomics. Based on the progress to date, it is apparent that structure will play an important role in the post-genomic era of biology.

Acknowledgment

We thank L. Zhang for producing the data in Box 2 and Fig. 1.

References

1 Gurd, F.R.N. and Rothgeb, T.M. (1979) Motions in proteins. *Adv. Protein Chem.* 33, 73–165
2 Laskowski, R.A. *et al.* (1996) X-SITE: use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins. *J. Mol. Biol.* 259, 175–201
3 Wallace, A.C. *et al.* (1996) Derivation of 3D coordinate templates for searching structural databases: application to Ser–His–Asp catalytic triads in the serine proteinases and lipases. *Protein Sci.* 5, 1001–1013
4 Henikoff, S. and Henikoff, J.G. (1991) Automated assembly of protein blocks for database searching. *Nucleic Acids Res.* 19, 6565–6572
5 Riley, M. (1993) Functions of gene products of *Escherichia coli*. *Microbiol. Rev.* 57, 862–952
6 Karp, P.D. and Riley, M. (1993) Representations of metabolic knowledge. *Ismb* 1, 207–215
7 Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
8 Pearson, W.R. (1996) Effective protein sequence comparison. *Methods Enzymol.* 266, 227–258
9 Sturrock, S.S. and Collins, J.F. (1993) *Biocomputing Research Unit*, University of Edinburgh, Edinburgh, UK
10 Bairoch, A. *et al.* (1995) The PROSITE database, its status in 1995. *Nucleic Acids Res.* 24, 189–196
11 Henikoff, S. and Henikoff, J.G. (1994) Protein family classification based on searching a database of blocks. *Genomics* 19, 97–107
12 Attwood, T.K. *et al.* (1994) PRINTS – A database of protein motif fingerprints. *Nucleic Acids Res.* 22, 3590–3596
13 Attwood, T.K. *et al.* (1997) Novel developments with the PRINTS protein fingerprint database. *Nucleic Acids Res.* 25, 212–216
14 Nevill-Manning, C.G. *et al.* (1998) Highly specific protein sequence motifs for genome analysis. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5865–5871
15 Fetrow, J.S. and Skolnick, J. (1998) Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and T1 ribonucleases. *J. Mol. Biol.* 281, 949–968
16 Yu, L. *et al.* (1998) A homology identification method that combines protein sequence and structure information. *Protein Sci.* 7, 2499–2510
17 Bork, P. and Bairoch, A. (1996) Go hunting in sequence databases but watch out for traps. *Trends Genet.* 12, 425–427
18 Gaasterland, T. (1998) Structural genomics: bioinformatics in the driver's seat. *Nat. Biotechnol.* 16, 625–627
19 McKusick, V.A. (1997) Genomics: structural and functional studies of genomes. *Genomics* 45, 244–249
20 Montelione, G.T. and Anderson, S. (1999) Structural genomics: keystone for a human proteome project. *Nat. Struct. Biol.* 6, 11–12

21 Fischer, D. *et al.* (1994) Three-dimensional, sequence order-independent structural comparison of a serine protease against the crystallographic database reveals active site similarities: potential implications to evolution and to protein folding. *Protein Sci.* 3, 769–778

22 Sanchez, R. and Sali, A. (1998) Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc. Natl. Acad. Sci. U. S. A.* 95, 13597–13602

23 Zhang, L. *et al.* (1998) Functional analysis of *E. coli* proteins for members of the α/β hydrolase family. *Fold. Design* 3, 535–548

24 Kasuya, A. and Thornton, J.M. (1999) Three-dimensional structure analysis of Prosite patterns. *J. Mol. Biol.* 286, 1673–1691

25 Coldren, C.D. *et al.* (1997) The rational design and construction of a cuboidal iron–sulfur protein. *Proc. Natl. Acad. Sci. U. S. A.* 94, 6635–6640

26 Pinto, A.L. *et al.* (1997) Construction of a catalytically active iron superoxide dismutase by rational protein design. *Proc. Natl. Acad. Sci. U. S. A.* 94, 5562–5567

27 Hellinga, H.W. and Richards, F.M. (1991) Construction of new ligand binding sites in proteins of known structure: (I) computer-aided modeling of sites with pre-defined geometry. *J. Mol. Biol.* 222, 763–785

28 Hellinga, H.W. *et al.* (1991) Construction of new ligand binding sites in proteins of known structure: (II) grafting of a buried transition metal binding site into *Escherichia coli* thioredoxin. *J. Mol. Biol.* 222, 787–803

29 Klemba, M. and Regan, L. (1995) Characterization of metal binding by a designed protein: single ligand substitutions at a tetrahedral Cys_2His_2 site. *Biochemistry* 34, 10094–10100

30 Klemba, M. *et al.* (1995) Novel metal-binding proteins by design. *Nat. Struct. Biol.* 2, 368–373

31 Farinas, E. and Regan, L. (1998) The *de novo* design of a rubredoxin-like Fe site. *Protein Sci.* 7, 1939–1946

32 Crowder, M.W. *et al.* (1995) Spectroscopic studies on the designed metal-binding sites of the 43C9 single chain antibody. *J. Am. Chem. Soc.* 117, 5627–5634

33 Halfon, S. and Craik, C.S. (1996) Regulation of proteolytic activity by engineered tridentate metal binding loops. *J. Am. Chem. Soc.* 118, 1227–1228

34 Wallace, A.C. *et al.* (1997) TESS: A geometric hashing algorithm for deriving 3D coordinate templates for searching structural databases: application to enzyme active sites. *Protein Sci.* 6, 2308–2323

35 Kleywegt, G.J. (1999) Recognition of spatial motifs in protein structures. *J. Mol. Biol.* 285, 1887–1897

36 Matsuo, Y. and Nishikawa, K. (1994) Protein structural similarities predicted by a sequence-structure compatibility method. *Protein Sci.* 3, 2055–2063

37 Russell, R.B. (1998) Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *J. Mol. Biol.* 279, 1211–1227

38 Han, K.F. *et al.* (1997) Three-dimensional structures and contexts associated with recurrent amino acid sequence patterns. *Protein Sci.* 6, 1587–1590

39 Artymiuk, P.J. *et al.* (1994) A graph-theoretic approach to the identification of three-dimensional patterns of amino acid side-chains in protein structures. *J. Mol. Biol.* 236, 327–344

40 Karlin, S. and Zhu, Z.Y. (1996) Characterizations of diverse residue clusters in protein three-dimensional structures. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8344–8349

41 Fetrow, J.S. *et al.* (1998) Functional analysis of the *Escherichia coli* genome using the sequence-to-structure-to-function paradigm: identification of proteins exhibiting the glutaredoxin/thioredoxin disulfide oxidoreductase activity. *J. Mol. Biol.* 282, 703–711

42 Abola, E.E. *et al.* (1987) *Protein Data Bank in Crystallographic Databases: Information Content, Software Systems, Scientific Application* (Allen, F.H. *et al.*, eds), Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester

43 Fetrow, J.S. *et al.* (1999) Structure-based functional motif identifies a potential disulfide oxidoreductase active site in the serine/threonine protein phosphatase-1 subfamily. *FASEB J.* 13, 1866–1874

44 Sali, A. *et al.* (1995) Evaluation of comparative protein modeling by MODELLER. *Proteins* 23, 318–326

45 Bystroff, C. and Baker, D. (1998) Prediction of local structure in proteins using a library of sequence-structure motifs. *J. Mol. Biol.* 281, 565–577

46 Shortle, D. (1999) The state of the art. *Curr. Biol.* 9, R205–R209

47 Lee, J. *et al.* (1999) Calculation of protein conformation by global optimization of a potential energy function. *Proteins* 3 (Suppl.), 204–208

48 Ortiz, A. *et al.* (1999) *Ab initio* folding of proteins using restraints derived from evolutionary information. *Proteins* 3 (Suppl.), 177–185

49 Bowie, J.U. *et al.* (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253, 164–170

50 Finkelstein, A.V. and Reva, B.A. (1991) A search for the most stable folds of protein chains. *Nature* 351, 497–499

51 Bryant, S.H. and Lawrence, C.E. (1993) An empirical energy function for threading protein sequence through folding motif. *Proteins* 16, 92–112

52 Lathrop, R. and Smith, T.F. (1996) Global optimum protein threading with gapped alignment and empirical pair scoring function. *J. Mol. Biol.* 255, 641–665

53 Jones, D.T. *et al* (1992). A new approach to protein fold recognition. *Nature* 358, 86–89

54 Sternberg, M.J. *et al.* (1999) Progress in protein structure prediction: assessment of CASP3. *Curr. Opin. Struct. Biol.* 9, 368–373

55 Miller, R.T. *et al.* (1996) Protein fold recognition by sequence threading tools and assessment techniques. *FASEB J.* 10, 171–178

56 Ortiz, A.R. *et al.* (1998) Fold assembly of small proteins using Monte Carlo simulations driven by restraints derived from multiple sequence alignments. *J. Mol. Biol.* 277, 419–448

57 Skolnick, J. *et al.* (1998) Reduced protein models and their application to the protein folding problem. *J. Biomol. Struct. Dyn.* 16, 381–396

58 Jaroszewski, L. *et al.* (1998) Fold prediction by a hierarchy of sequence, threading and modeling methods. *Protein Sci.* 7, 1431–1440

59 Takahashi, M. *et al.* (1996) Locations of functional domains in the RecA protein: overlap of domains and regulation of activities. *Eur. J. Biochem.* 242, 20–28

60 Leong, L.E. *et al.* (1993) Human rhinovirus-14 protease 3C (3Cpro) binds specifically to the 5′ noncoding region of the viral RNA: evidence that 3Cpro has different domains for the RNA binding and proteolytic activities. *J. Biol. Chem.* 268, 25735–25739

61 Matthews, D.A. *et al.* (1994) Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site and means for cleaving precursor polyprotein. *Cell* 77, 761–771

62 Ladomery, M. (1997) Multifunctional proteins suggest connections between transcriptional and post-transcriptional processes. *BioEssays* 19, 903–909

63 Goldberg, J. *et al.* (1995) Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* 376, 745–753

64 Mumby, M.C. and Walter, G. (1993) Protein serine/threonine phosphatases: structure, regulation and functions in cell growth. *Physiol. Rev.* 73, 673–699

65 Jia, Z. (1997) Protein phosphatases: structures and implications. *Biochem. Cell Biol.* 75, 17–26

66 Holmes, C.F.B. and Boland, M.P. (1993) Inhibitors of protein phosphatase-1 and -2A: two of the major serine/threonine protein phosphatases involved in cellular regulation. *Curr. Opin. Struct. Biol.* 3, 934–943

67 Nemani, R. and Lee, E.Y.C. (1993) Reactivity of sulfhydryl groups of the catalytic subunits of rabbit skeletal muscle protein phosphatases 1 and 2A. *Arch. Biochem. Biophys.* 300, 24–29

68 Murzin, A.G. *et al.* (1995) Scop: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540

69 Orengo, C.A. *et al.* (1997) CATH: a hierarchic classification of protein domain structures. *Structure* 5, 1093–1108

70 Lesk, A.M. *et al.* (1989) Structural principles of α/β proteins: the packing of the interior of the sheet. *Proteins Struct. Funct. Genet.* 5, 139–148

71 Farber, G.K. and Petsko, G.A. (1990) The evolution of α/β barrel enzymes. *Trends Biochem. Sci.* 15, 228–234

72 Hegyi, H. and Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol.* 288, 147–164

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

Current Understanding of Food Allergens

SAMUEL B. LEHRER, ROSALIA AYUSO, AND GERALD REESE

Section of Clinical Immunology, Allergy and Rheumatology, Department of Medicine, Tulane University School of Medicine, New Orleans, Louisiana 70112, USA

ABSTRACT: **Food allergies are IgE-mediated immunological reactions; this distinguishes them from other adverse reactions to foods. Most (>90%) of the recognized food allergies are generally thought to be caused by eight foods or food groups. A number of factors can affect food allergy development, including diet and culture, route of exposure, processing, cooking, and digestion. In addition, it is thought that the properties of certain food proteins render them more likely to be allergenic than other proteins. Most food allergens are major proteins, polyvalent molecules with at least two or more IgE-binding sites, and are recognized as foreign molecules (hence immunogenic). A number of major food allergens have been recently characterized, and amino acid sequences determined. Tropomyosin is the only major allergen of shrimp. A number of IgE-binding epitopes have been identified in this molecule, though they may vary from one shrimp-allergic individual to another. Single amino acid substitutions within epitopes based on that of homologous, nonreactive tropomyosins can substantially enhance or abolish IgE antibody binding. Using the accumulated knowledge of food allergen protein structure, the allergenicity of novel proteins to which there has been no prior human exposure has been assessed. This has been based primarily on the lability or resistance of a protein to enzymatic degradation. Clearly, further criteria must be developed to refine this process. In this regard, the development of animal models that have been sufficiently validated as surrogates of human IgE antibody responses is needed for more precise assessment of the allergenic potential of proteins.**

KEYWORDS: **food allergens; metabolic food intolerance; gut-associated lymphoid tissue; shrimp tropomyosin**

Address for correspondence: Samuel Lehrer, Ph.D., Research Professor of Medicine, Tulane University Medical Center School of Medicine, Department of Medicine SL57, Section of Clinical Immunology and Allergy, 1700 Perdido Street, New Orleans, LA 70112-1293. Voice: 504-588-5578; fax: 504-584-3686.
sblehrer@tulane.edu

Ann. N.Y. Acad. Sci. 964: 69–85 (2002).

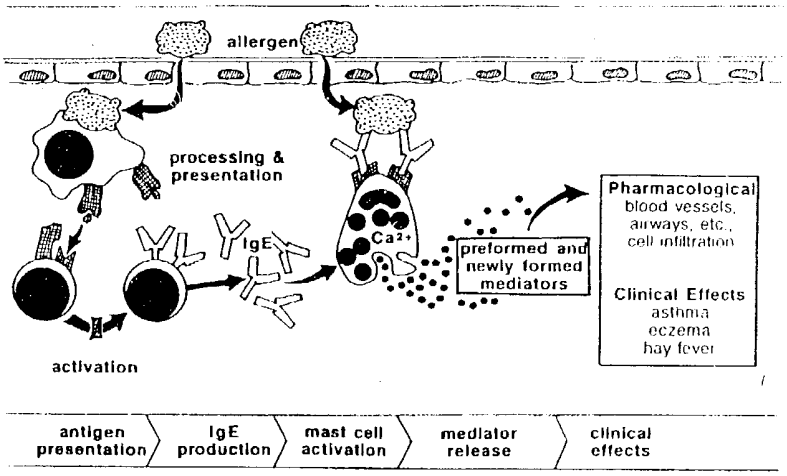


FIGURE 1. Induction and effective mechanisms in IgE-mediated hypersensitivity: An allergen that crosses mucosal membrane barriers is processed and presented to IgE-producing cells (antigen presentation). This stimulates the production of IgE antibodies. These antibodies have the ability to bind to mast cell surfaces (mast cell activation); subsequent interaction with allergen can cross-link two or more membrane-bound IgE molecules, which stimulates the release of preformed and newly synthesized mediators of allergy. These mediators result in the clinical effects of allergic reactions.

INTRODUCTION

It has been estimated that approximately 1 to 2% of the adult population and 4 to 8% of the pediatric population suffer from food allergies.[1–3] Most of the recognized food allergies (more than 90%) are generally thought to be caused by eight foods or food groups. These include crustacea, eggs, fish, milk, peanuts, soybeans, tree nuts, and wheat. The symptoms of food allergy vary considerably from mild discomfort, such as itching of the skin or throat, to profound, life-threatening, systemic anaphylactic shock. Food allergies are IgE-mediated immunological reactions, which distinguishes them from other adverse reactions to foods that may include metabolic food intolerances or toxic food reactions.[3]

The induction of allergic reactions in general, and food allergic reactions specifically, is summarized in FIGURE 1. An allergen, or an immunologically intact fragment of allergen resistant to cooking and digestive processes, crosses the epithelial barrier of the gut. After processing and presentation by

antigen-presenting cells (generally macrophages), immunologically active fragments are recognized by T lymphocytes, which along with other allergen fragments stimulate B lymphocytes to synthesize and ultimately produce IgE antibodies specific for that allergen of stimulation. IgE antibodies have the unique ability to fix to the surfaces of mast cells or basophils. Upon a second or subsequent interaction of allergen with cell-bound IgE molecules, events are initiated in the membrane of these cells leading to the release of both preformed and newly formed mediators. These mediators are molecules that have potent pharmacological reactions resulting in the clinical effects that are seen in allergic reactions.[1,2] When one views the events leading to the induction and manifestation of allergic reaction, there are clearly certain features imposed on these allergenic molecules by their biological activity. For example, food allergens must be rather hearty molecules to resist the processes of cooking and digestion; they must be small enough to cross the epithelial barrier of the gut, yet of sufficient size to remain in an immunologically intact form. Finally, these molecules must be at least divalent and most probably polyvalent to cross-link cell-bound IgE molecules resulting in allergic reactions.[4–6]

There are a number of factors that can affect food allergy development. Diet and culture can have a significant effect on food allergies that manifest themselves in a particular region. For example, codfish allergy is most prevalent in Norway because codfish is an important source of protein for Norwegians. Similarly, rice and soy allergy are more prevalent in Japan, and peanut allergy is more prevalent in the United States. The route of exposure can also have an important effect on food-induced allergic reactions. Clearly, ingestion is a much harsher route of allergen exposure and thus may require higher doses in order for a sufficient number of molecules or allergenically active molecular fragments to survive in an immunologically intact form. However, individuals can also be sensitized through inhalation of allergen, a less harsh route in which lower doses of allergen can stimulate a response.[7] This occurs with occupational sensitization of food workers.[8] The relationship of inhalation to ingestion exposures of food allergen or the cross-reactivity of food allergens with inhaled allergens in the pathogenesis of food allergy is only now beginning to be understood.

Processing, cooking, and digestion obviously can have a profound effect on foods.[4–6] Because of the rather harsh treatments to which they are exposed, food allergens are generally thought to be more resistant to these processes than other antigens. This is not true of all food allergens, however; and, clearly, food allergens have been identified that are not resistant to these processes. Finally, one must consider intrinsic factors present in foods that seem to affect food allergen development. Why are certain molecules apparently much more allergenic than others? Is this only due to their resistance to various cooking and digestive processes? Clearly, many nonallergenic molecules are equally resistant to food processing methods or digestion, and thus it is

thought that there are other factors present in these food molecules that make them more allergenic. Further studies are clearly needed to elucidate these factors.

Thus, the investigation of food allergens presents unique challenges when compared to studies of other allergens. Exposure to foods is more complicated than that to other antigens in that most food antigen exposure occurs through ingestion, a route that can significantly alter food molecules, resulting in molecules or molecular fragments substantially different from those that are identified in food extracts. It is clear that absorption through the gut is needed to induce a food-allergic response, and thus the ability of molecules to be absorbed through the gut and the role of gut-associated lymphoid tissue, which may be different from that of the central immunological tissue, need to be better understood. The effects on the induction of food allergies of inhalation compared to ingestion of food allergens needs to be better understood. Foods are generally complex, consisting of over 10,000 to 20,000 protein molecules and thus present more of a challenge for study. Finally, it is very difficult, if not impossible, to assess the allergen dose responsible for either sensitizing or provoking a food-allergic response or identify molecular structure of active fragments. Some information is available through food-challenge studies on the amount of food that provokes allergic reactions, but unfortunately nothing is known on sensitizing doses. This, again, is due to the complexity of the induction of food-allergic reactions.[3]

PROPERTIES OF FOOD ALLERGENS

Although the precise properties that render a food molecule allergenic are unknown, some general characteristics of allergenic food proteins have been identified. Generally, many major food allergens are major food proteins. This suggests that higher doses of a particular allergen enhance the likelihood of producing an allergic response. The fact that most food allergens are usually expressed as percent of total food protein can confuse the issue of exposure doses. For example, when using percent protein as a measurement of a molecule, low-protein foods such as potatoes may contain a protein such as patatin that is 50% of the potato's total protein yet is only a minor potato component (1 g/100 g dry weight), whereas a protein of a similar amount in high-protein foods constitutes a much lower percentage (2–4%) of the total food protein. Clearly, one must take into account the amount of a protein consumed when considering exposure. Food allergens are molecules that must be capable of stimulating immune responses; this is an absolute requirement because molecules that are unable to stimulate antibody production cannot induce an immediate hypersensitivity response. Generally, food allergens are very stable molecules that resist the effects of food processing, cooking, and

digestive enzymes in the gut; intellectually, this makes sense in that molecules that have such abilities would more likely be allergenic. This is not an absolute requirement, however, and there are stable food molecules that are not allergenic, while conversely, several food allergens have been identified that are not highly resistant to these processes.

Food allergens must be polyvalent molecules with at least two or more IgE antibody binding sites. This is because the mast cell or basophil-bound IgE must be cross-linked by these allergen molecules in order to stimulate mediator release. This is an absolute requirement, although one cannot preclude monovalent molecules aggregating *in vivo*. Generally, most food allergens are between 10 and 70 kDa and are primarily glycoproteins with acid isoelectric points; however, this is also the case for most antigens, and thus it is not a distinguishing feature for food allergens.[2–4]

The composition of allergenic foods in comparison to nonallergenic foods has always been of interest. The aim, of course, is to be able to identify differences that may be used to identify foods more likely to be allergenic. Comparison of the composition of common allergenic and nonallergenic foods in our diet is summarized in TABLE 1. In general, high-protein foods are more allergenic; of the 23 foods that contain >20% protein, 14 (61%) have been identified as important food allergens. Of the foods that contain <20% protein, only 5/15 (33%) are important allergenic foods. It is interesting to note that of the eight allergenic foods with somewhat lower protein concentration (<30%), 5/8 (63%) have high lipid concentrations (>50%) in comparison to other components; could elevated lipid levels enhance the allergenicity of these proteins? Interestingly, almost all low-protein foods (<20%) have very high concentrations of carbohydrates, with the exception of the tree nuts. Of the 15 foods with <20% protein concentration listed in this table, (omitting tree nuts which have high lipid concentrations) only one (wheat) is considered a major food allergen, and the remaining fruits and vegetables have some allergenic activity but are not major food allergens.

What is the allergen concentration or dose that is necessary to stimulate sensitization to a particular allergen or indeed provoke an allergic reaction? Unfortunately, there is no direct way this can be determined without developing relevant animal models. Perhaps such information could be indirectly obtained by estimating the dose of allergen to which individuals are exposed in relationship to their allergic responses or IgE antibody production (TABLE 2). For this reason, it is very important to express particular protein amounts in terms of dose rather than percent protein, which, as discussed previously, can be misleading.

Cross-reactivity among a number of foods of similar food types as well as between foods and seemingly unrelated substances has been demonstrated. For many years it has been known that foods within a certain group or family can cross-react; examples are members of the legume family, such as soybeans, peanuts, and peas,[9,10] and crustaceans, such as lobster, shrimp, and

TABLE 1. Composition of common foods in our diet

Food 850.5 g (3 ozs)	Total protein g (%)[a]	Total fat g (%)[a]	Total carbohydrate g (%)[a]	Water g (%) water)
Cod, Atlantic fillet, baked	19.39 (96.4)	0.73 (3.6)	0 (0)	64.55 (75.90)
Tuna, light, canned	24.52 (96.1)	1.04 (3.9)	0 (0)	53.4 (62.78)
Shrimp, small, mixed species, steamed	17.78 (95.1)	0.92 (4.9)	0 (0)	65.74 (77.30)
Crawfish, mixed species, boiled	14.88 (93.1)	1.11 (6.9)	0 (0)	68.72 (66.63)
Crab, blue, steamed	17.18 (91.9)	1.51 (8.1)	0 (0)	65.83 (77.40)
Lobster, northern steamed	17.44 (91.6)	0.50 (2.6)	1.09 (5.7)	64.64 (76.00)
Crab, Dungeness crab, steamed	18.97 (91.1)	1.05 (5.0)	0.81 (3.9)	62.34 (66.63)
Salmon, fillet, pink, baked	21.77 (85.3)	3.76 (14.7)	0 (0)	59.28 (69.70)
Chicken breast, boneless, roasted	25.34 (79.3)	6.62 (20.7)	0 (0)	53.07 (62.40)
Tuna, canned	24.75 (78.0)	6.98 (22.0)	0 (0)	50.86 (69.70)
Beef, roast, lean only	23.97 (76.8)	7.23 (23.2)	0 (0)	52.70 (61.63)
Trout, fillet, mixed species, baked	22.62 (75.9)	7.20 (24.2)	0 (0)	53.92 (63.40)
Pork, fresh whole loin, roasted	23.05 (64.8)	12.50 (35.2)	0 (0)	48.90 (57.50)
Lamb, NZ leg, roasted 1/8" trim	21.52 (64.4)	11.91 (35.6)	0 (0)	50.18 (59.0)
Egg	10.63 (52.7)	8.51 (42.2)	1.04 (5.2)	64.04 (75.3)
Soy beverage, fluid	2.34 (42.5)	1.62 (29.5)	1.54 (28.0)	79.35 (93.30)
Soy flour, whole AMI	30.24 (39.0)	30.24 (39.0)	17.01 (22.0)	4.39 (5.16)
Broccoli, boiled	2.64 (36.2)	0.10 (1.4)	4.55 (62.4)	77.14 (90.70)
Milk, whole, cow's*	3.01 (29.1)	3.06 (29.6)	4.26 (41.2)	80.52 (88.0)
Peanuts, dry roasted	20.16 (25.0)	42.27 (52.4)	18.29 (22.7)	1.32 (1.55)
Almonds, dry roasted	18.12 (22.8)	45.25 (56.8)	16.24 (20.4)	1.79 (2.10)
Okra, boiled	1.59 (20.2)	0.14 (1.8)	6.13 (78.0)	76.45 (89.90)
Stringbeans, canned	0.98 (16.4)	0.09 (1.5)	3.83 (64.2)	79.35 (93.30)
Wheat grain, durum	11.65 (15.7)	2.10 (2.8)	60.47 (81.5)	9.27 (10.90)
Brazil nuts, large dried	12.16 (15.3)	56.30 (71.0)	10.89 (13.7)	2.84 (3.33)
English walnuts, dried	12.16 (15.1)	52.65 (65.5)	15.56 (19.4)	3.10 (3.69)
Corn, yellow frozen	2.34 (12.1)	0.37 (1.9)	16.67 (86.0)	65.23 (76.7)
Apple	0.16 (11.9)	0.31 (2.3)	13.1 (96.5)	71.36 (83.90)
Flour, all purpose white enriched	8.79 (11.8)	0.83 (1.1)	64.89 (87.1)	10.12 (11.90)
Eggplant, boiled	0.71 (10.8)	0.20 (3.0)	5.65 (86.1)	78.08 (91.80)
Carrots, boiled	0.93 (9.3)	0.15 (1.5)	8.93 (89.2)	74.33 (87.40)
Pecans, dried halves	6.59 (8.3)	57.49 (72.3)	15.48 (19.5)	4.10 (4.82)
Corn flour, whole grain white	5.89 (7.9)	3.28 (4.4)	65.40 (87.7)	9.27 (10.90)
Potato flour	5.87 (7.6)	0.29 (0.4)	70.68 (92.0)	5.55 (6.53)
Orange	0.80 (7.3)	0.10 (0.9)	10.04 (91.8)	73.82 (86.80)
Banana	0.88 (4.2)	0.41 (1.9)	19.90 (93.9)	63.19 (74.30)
Grapes, Thompson seedless	0.56 (3.5)	0.49 (3.0)	15.14 (93.5)	68.55 (80.60)

[a] Percent dry weight.
[b] Percent total weight.

crayfish.[11,12] However, these cross-reactivities may have very different clinical endpoints. Cross-reactivity among legumes demonstrated by *in vitro* tests is not necessarily clinically significant.[13] However, cross-reactivity among crustacea can be clinically significant, and certainly allergy to one crustacean suggests potential reactivity to other crustacea, although not absolute. Cross-

TABLE 2. Food allergen concentration

Protein	% Total protein
Egg white allergens	
Ovalbumin (Gal d 2)	54
Ovomucoid (Gal d 1)	11
Conalbumin (Gal d 3)	12–13
Ovomucin	1.5–3.5
Lysozyme	3.4–3.5
Milk allergens	
β-Lactoglobulin	9
Casein	80
BSA	1
α-Lactalbumin	4
Soybean allergens	
β-Conglycinin (β subunit)	18.5
Kunitz trypsin inhibitor	2–4
Soy lectin	1–2
β-Conglycinin (α subunit)	18.5
Glycinin	51
Gly m Bd 30K	2–3
Peanut allergens	
Vicilin (Ara h 1)	1[a]
Conglutin family (Ara h 2)	6[b]
Peanut lectin	1.3
Mustard allergens	
Sin a 1	20
Bra j 1	20
Fish allergens	
Parvalbumin (Gad c 1)	0.22–0.44
Shrimp allergens	
Tropomyosin (Pen a 1)	20[c]
Potato allergens	
Patatin (Sol t 1)	30–40

[a]Crude peanut meal.
[b]Crude peanut extract.
[c]Crude shrimp extract.

reactivity can also occur between foods and seemingly unrelated substances, and indeed this has been of greater interest in recent years. Cross-reactivity has been demonstrated with different pollens, particularly birch pollen, with a variety of fruits and vegetables, based on investigations reported primarily in Europe.[14–17] More recently, with increased interest in latex allergy, cross-reactivity has also been demonstrated between latex proteins and those present in fruits and vegetables.[18,19] Indeed, these cross-reactivities of fruits and/or vegetables with inhaled allergens have clinical significance in that some patients have developed food allergies subsequent to inhalant exposure. It has been shown by *in vitro* assays that crustacea and mollusks can cross-

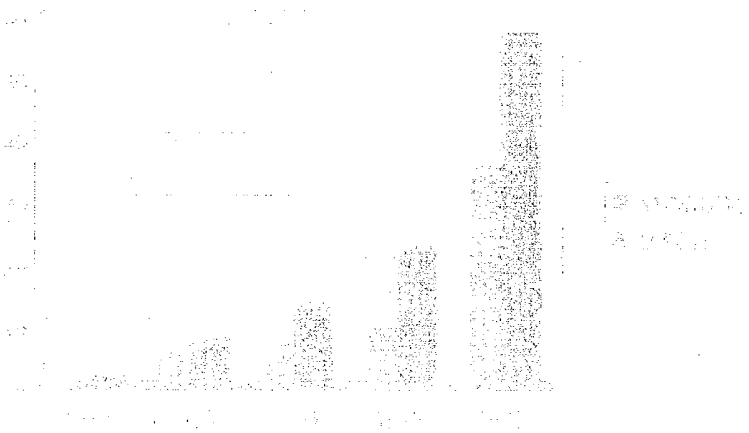


FIGURE 2. Common food allergens identified, sequenced, and cloned from 1980 through 2000. The number of food allergens cloned has increased substantially during the past 20 years, and the results to date suggest that it will continue to increase at a very rapid pace. As of 2000, 57 food allergens had been cloned.

react,[20] probably due to the muscle protein tropomyosin, a common allergen present in crustacea. Whether this cross-reactivity is clinically significant or not remains to be determined, since many crustacea-allergic individuals can consume mollusks without reactivity. Indeed, classic IgE-mediated allergic reactions to mollusks with several exceptions[21–23] have not been well described.

Since 1980, a number of food allergens have been identified and sequenced (FIG. 2). Although common structural properties have been sought, no unique structural feature for food allergen has been identified. Over the years, the number of food allergens identified and characterized have substantially increased. To date, the amino acid sequences of at least 57 food allergens have been determined; based on available technologies, this number will increase substantially over the next several years. More information on allergen structure will provide a more accurate allergenicity assessment, particularly since one approach to characterization of molecules of unknown allergenicity is based on amino acid sequence similarities.[4,6]

FOOD ALLERGEN EPITOPES: SHRIMP TROPOMYOSIN

The major allergen in shrimp and indeed other crustacea has been shown to be the muscle protein tropomyosin.[24–28] Tropomyosin is a major (approx-

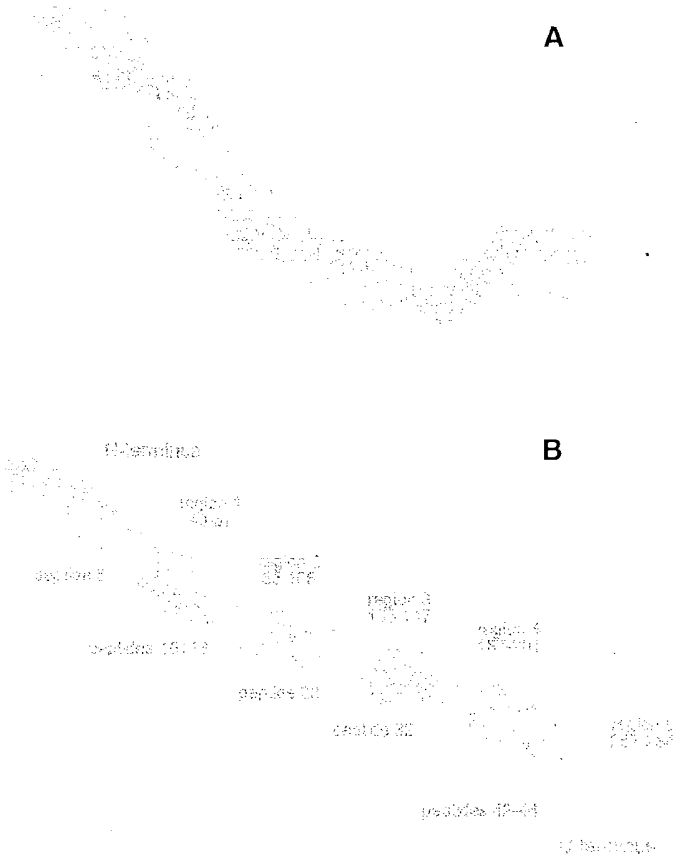


FIGURE 3. Coiled-coil structure of the tropomyosin molecule **(top, A)** and this structure with IgE-binding regions identified **(bottom, B)**. Tropomyosin molecule is represented schematically as two chains (red and blue) that are in alpha-helical formation, each chain wound around the other, demonstrating the typical coiled-coil structure of tropomyosin. The structure is typical for all tropomyosin molecules. The same molecule is shown in the bottom portion of the figure **(B)** with the IgE-binding regions highlighted in yellow. Each chain of tropomyosin is identical and parallel. IgE-binding regions appear to exist throughout the entire region portion of the molecule.

imately 20% of the dry weight) shrimp protein, is the only major allergen identified in shrimp, and is thought to be a major allergen in other invertebrates as well.[29–32] Tropomyosin molecules consist of two identical polypeptide chains, each in an alpha-helical formation coiled around the other; this is

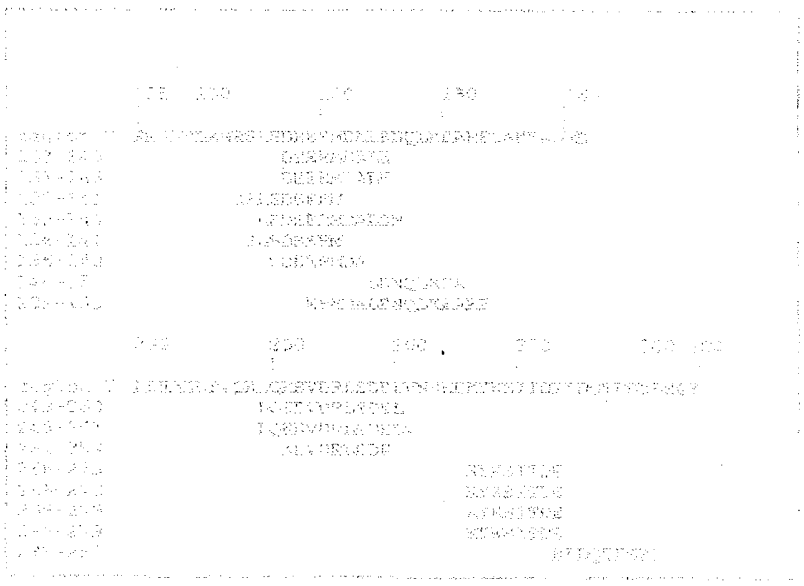
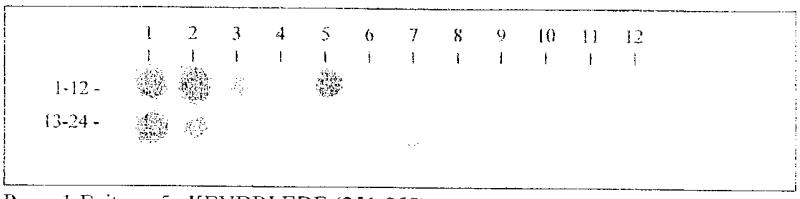
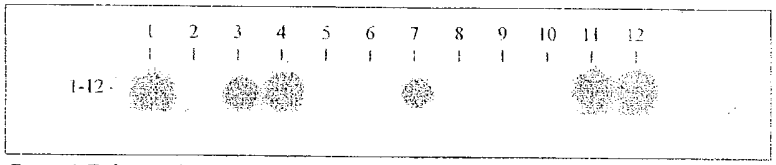


FIGURE 4. Individually recognized epitopes of regions 3 and 5 of the major shrimp allergen, Pen a 1. Epitopes, recognized by individual region-reactive sera, are summarized in the figure. It is interesting to note that the six sera reactive with epitope 3a appear to generally vary considerably in their reactivity, all reacting to a common core of five amino acids. Variability is seen with region 3b as well. In contrast, the variability appears to be reduced for individual sera reacting to epitopes 5a and 5b.

referred to as the coiled-coiled structure of tropomyosin,[33] as illustrated in FIGURE 3. Studies in our laboratory[34] have identified five major IgE-binding regions present in shrimp tropomyosin (FIG. 3). The IgE-binding epitopes within these regions have been characterized using synthetic overlapping peptides of 5–15 amino acids in length with an offset of one to two amino acids for a particular region. Region-specific antisera were tested for IgE antibody reactivity, and an epitope was identified as the shortest peptide with maximal IgE binding to that antiserum. Using these results, eight IgE-binding epitopes in five main allergenic regions have been detected.[35] An example of epitopes identified in two major regions, region 3 and region 5, is shown in FIGURE 4. Of interest is the fact that there can be significant variability of individual serum reactivity to epitopes. For example, reactivity to epitope 3a demonstrates significant wobble in IgE antibody reactivity with a common core region of five amino acid residues being the same. In epitopes 5a and 5b, the wobble effect is much reduced, and reactivity of individual sera is fairly



Pen a 1 Epitope 5a:KEVDRLEDE (251-259)



Pen a 1 Epitope 5b:KYKSITDE (266-273)

FIGURE 5. Combinatorial substitutions in epitopes 5a and 5b. **Pen a 1 Epitope 5a:** KEVDRLEDGE (251–259) (spot 1); substitution 2 (S for E at position 252), substitution 3 (I for V at position 253), substitution 4 (D for R at position 255), substitution 5 (S for E and I for V), substitution 6 (S for E and D for R), substitution 7 (I for V and D for R), substitution 8 (S for E, I for V, and D for R), substitution 13 (T for E at position 252), substitution 14 (T for E and I for V), substitution 15 (T for E and D for R), and substitution 16 (T for E, I for V, and D for R). **Pen a 1 Epitope 5b:** KYKSITDE (266–273) (spots 1 and 11); substitution 2 (F for S at position 269), substitution 3 (C for T at position 271), substitution 4 (D for E at position 273), substitution 5 (F for S and C for T), substitution 6 (F for S and D for E), substitution 7 (C for T and D for E), substitution 8 (F for S, C for T, and D for E), and substitution 12 (S for T at position 271).

similar. Perhaps this reflects differences in affinity of IgE antibodies to different epitopes and raises the question whether IgE antibody reactivity to certain epitopes may be more relevant clinically than that to other epitopes on the same allergen.

To identify amino acid residues in an epitope essential for IgE antibody binding, mutated peptide epitopes with substituted amino acids based on the amino acid sequence of homologous regions in nonallergenic tropomyosins were synthesized and tested for IgE antibody binding with epitope residue antisera. An example of the IgE antibody binding activity to mutated epitopes 5a and 5b is shown in FIGURE 5. The results indicate that one amino acid substitution may have no effect, reduced or abolished IgE binding, or even enhanced IgE binding.[35–36] This certainly has significant implications when assessing IgE antibody reactivity to food allergen epitopes. Generally, greater than two substitutions usually abolish IgE-binding ability to modified Pen a 1 peptides. Substitutions from the epitope center are more likely to eliminate (59.5% IgE

binding) as opposed to substitutions on the epitope periphery (39.1%). Most (63.5%) of the single substitutions that eliminate IgE binding are not conservative (different amino acid groups) as compared to the 23.1% of the conservative (same amino acid group) substitutions.[35] Whether these observations for IgE binding epitopes of tropomyosin are relevant to those of other food allergens is not clear. Generally, however, our results are consistent with those reported for epitopes of major peanut allergens Ara h 1, 2, and 3.[37–39] Common patterns of epitope structure and activity observed for food allergens would provide important information for assessment of the allergenicity of food proteins.

PREDICTING FOOD PROTEIN ALLERGENICITY: FOOD FOR THOUGHT

Based on current knowledge of the properties of known food allergens, can the allergenicity of a novel protein to which there is no prior human exposure be precisely predicted? The answer at this time is probably no; however, reasonable assessments based on our current knowledge of food allergens can be made. The information available on known food allergens is increasing, and thus the precision in predicting food protein allergenicity will increase as well. Several important issues are relevant to any type of predictive model. First, food allergen concentration in relationship to food sensitization or provocation of an allergic reaction needs to be determined. Unfortunately, little, if any, information is available on the sensitizing dose of a food allergen other than crude estimates of doses in prior exposures; there is more information, although sparse, on the allergen dose provoking an allergic reaction. Better estimates of exposure levels to known food allergens that stimulate IgE antibody production or provoke allergic reactions will aid development of an indirect measurement of sensitization dose. If sensitizing doses for different food allergens can be estimated and sensitization does not occur below such doses, can a level of food allergens be determined below which sensitization does not occur? Such information, obtained through analysis of the doses to which we are exposed to food allergens, could be invaluable in assessing novel protein allergenicity.

Second, a protein's ability to resist degradation by processing and enzymatic digestion has been used to assess potential allergenic and nonallergenic food molecules; indeed, it has been a criterion for assessing the potential allergenicity of novel proteins. Nevertheless, many food proteins that are not allergenic are stable to processing and enzymatic digestion. Conversely, some protein molecules have been shown to be allergenic yet are sensitive to pro-

cessing and/or enzymatic digestion. Does this mean that stability of molecules is not important in allergenicity assessment? Probably not, but it does tell us that this is not an absolute criterion, and one must be careful in using this information for assessing proteins of unknown allergenicity.

Additional criteria to assess the allergenicity of proteins must be developed. The sensitizing dose for an allergen mentioned previously may be useful in estimating the amount of a food allergen component below which the induction of an allergic response is unlikely. Many high-protein foods such as shellfish are major food allergens; however, there are other high-protein foods that contain similar molecules, such as meats like beef or chicken, that are not very allergenic. These proteins from less allergenic sources are more similar to molecules expressed in man. Thus, the foreignness (immunogenicity) of certain proteins has an effect on allergenicity. For example, crustacea belong to a different phylum than humans, cattle, or pigs; and thus crustacean tropomyosins are more foreign and therefore more allergenic in humans compared to the more closely related beef or pork tropomyosins.

Amino acid sequence similarity is used to compare protein molecules with known allergens; however, the criteria for comparison have not been adequately described. What characteristics are important to compare? Are selected sequence comparison of epitopes (B or T cell) useful or are comparisons of the entire molecule more significant? What length of amino acid sequence and what percent of identity are necessary for a match? Such criteria are particularly important since it has recently been shown that small sequence changes, even in one amino acid, of IgE-binding epitopes in peanut or shrimp allergens may abolish or even enhance IgE antibody reactivity.[35–39] Thus, under current assessment methods, theoretically an amino acid sequence from a protein of unknown allergenicity may have 90% identity with an allergenic epitope, yet according to current criteria may not be included as a potential risk. Alternatively, should amino acid sequence homology of novel proteins with nonallergens be considered as important? If a protein has a sequence homology of more than 80% with a known nonallergen, does that fact rate importance in allergenicity assessment? Should the similarity or dissimilarity to human proteins (i.e. foreignness) be assessed? All of these issues need further consideration.

Finally, development of animal models that have been adequately validated as surrogates of human IgE antibody responses is required for more precise assessments of the allergenic potential of proteins. Such models must be adequately described with regard to the type of exposure, the route of exposure, the number of exposures, the allergen dose, and the use of adjuvants. My own prejudice is that an animal model that reflects all aspects of food allergy is not only unlikely to be developed, but also probably not necessary for assessment of novel protein allergenicity. However, an animal model of allergenicity, although reflecting only one aspect of food allergy, would serve to provide an important basis for testing or assessing the allergenicity of novel

proteins. For validation, animal models must react to allergenic foods as opposed to nonallergenic foods and recognize protein allergens within the allergenic foods as opposed to nonallergenic proteins within the same foods. This can be easily assessed by Western blotting with animal serum compared to human sera from food-allergic individuals. Finally, epitope reactivity of animal IgE antibodies must be compared to that of humans. One would expect that a validated animal model of allergenicity would produce IgE antibodies that react to epitopes similarly to human IgE antibodies. If all of these criteria correlate with human allergic responses, such a model will be an invaluable tool for the assessment of not only novel proteins but any new proteins to which persons are exposed.

Clearly, our body of knowledge of food allergen structure has increased substantially during the past 20 years. Allergens are molecules that generally (although not always) appear to be major protein components within a particular food and are usually resistant to processing, cooking, and digestion. Further study of the general properties of foods in relationship to their allergenicity is encouraged in that this may provide more useful information on the basis of allergenicity of these molecules. For example, foods of medium protein concentration that are highly allergenic, such as peanuts and tree nuts, appear to have higher levels of fat. This might be useful in protecting allergenic molecules during the digestive process. Furthermore, little information is known about the dose of allergen required to stimulate immune response. Little consideration has been given to homology of allergens in relationship to their foreignness. These are certainly important areas that warrant further study. The evidence to date suggests that we cannot yet precisely predict the potential allergenicity of a molecule from which there is no prior human exposure. However, based on our knowledge of food allergens, it is possible to make educated estimations about the potential allergenicity that should exclude any major new allergens from being introduced into our food supply through genetic modification of foods.

ACKNOWLEDGMENTS

The authors wish to acknowledge the excellent help in preparation of this manuscript by Ms. Pat Constant and the information provided on the composition of commonly consumed foods by Dr. Carol O'Neil at the Department of Nutrition, Louisiana State University in Baton Rouge, Louisiana. Funds provided by the National Fisheries Institute and the Department of Medicine, Tulane University School of Medicine, helped cover the cost of preparation of this manuscript.

REFERENCES

1. BURKS, A.W. & H.A. SAMPSON. 1993. Food allergies in children. Curr. Prob. Pediatr. **23:** 230–252.
2. JANSEN, J.J., A.F. KARDINAAL, G.H. HUIJBERS, *et al.* 1994. Prevalence of food allergy and intolerance in the adult Dutch population. J. Allergy Clin. Immunol. **93:** 446–456.
3. SAMPSON, H.A. & D.D. METCALFE. 1991. Immediate reactions to foods. *In* Food Allergy: Adverse Reactions to Foods and Food Additives. D.D. Metcalfe, H.A. Sampson & R.A. Simon, Eds.: 99–112. Blackwell Scientific Publications. Oxford.
4. LEHRER, S.B. & G. REESE. 1998. Food allergens: implications for biotechnology. *In* Biotechnology and Safety Assessment, 2nd ed.: 127–149. Taylor & Francis. Philadelphia.
5. TAYLOR, S.L. & S.B. LEHRER. 1996. Principles and characteristics of food allergens. Crit. Rev. Food Sci. Nutr. **36(S):** S91–S118.
6. LEHRER, S.B., W.E. HORNER & G. REESE. 1996. Why are some proteins allergenic? Implications for biotechnology. Crit. Rev. Food Sci. Nutr. **36(6):** 553–564.
7. NELDE, A., M. TEUFEL, C. HAHN, *et al.* 2001. The impact of the route and frequency of antigen exposure on the IgE response in allergy. Intl. Arch. Allergy Immunol. **124:** 461–469.
8. O'NEIL, C.E. & S.B. LEHRER. 1996. Occupational reactions to food allergens. *In* Food Allergy: Adverse Reactions to Foods and Food Additives. D.D. Metcalfe, H.A. Sampson & R.A. Simon, Eds.: 311–335. Blackwell Scientific Publications. Oxford.
9. BARNETT, D., B. BONHAN & W.E.H. HOWDEN. 1987. Allergenic cross-reactions among legume foods—in vitro study. J. Allergy Clin. Immunol. **79:** 443–448.
10. TAYLOR, S.L., R.F. LEMANSKI, R.K. BUSH & W.W. BUSSE. 1987. Food allergens: structure and immunologic properties. Ann. Allergy **59:** 93–99.
11. DAUL, C.B., J.E. MORGAN, N.P. WARING, *et al.* 1987. Immunological evaluation of shrimp-allergic individuals. J. Allergy Clin. Immunol. **80:** 716–722.
12. LEHRER, S.B. 1986. The complex nature of food antigens: studies of cross-reacting crustacea allergens. Ann. Allergy **57:** 267–272.
13. BERNHISEL-BROADBENT, J. & H.A. SAMPSON. 1989. Cross-allergenicity in the legume botanical family in children with food hypersensitivity. J. Allergy Clin. Immunol. **83:** 435–440.
14. ANDERSON, L.B., E.M. DREYFUSS, J. LOGAN & D.E. JOHNSTON. 1970. Melon and banana sensitivity coincident with ragweed pollinosis. J. Allergy Clin. Immunol. **45:** 310–319.
15. ENBERG, R.N., F.E. LEICKLEY, J. MCCOULLOUGH, *et al.* 1987. Watermelon and ragweed share allergens. J. Allergy Clin. Immunol. **79:** 867–875.
16. CALKHOVEN, P.G., M. AALBERS, V.L. KOSHTE, *et al.* 1987. Cross-reactivity among birch pollen, vegetables and fruits as detected by IgE antibodies is due to at least three distinct cross-reactive structures. Allergy **42:** 382–390.
17. WÜTRICH, B. & R. DIETSCHI. 1985. The celery-carrot-mugwort-condiment syndrome: skin test and RAST results. Shweiz Med. Wochenschr. **115:** 258–264.

18. MÄKINEN-KILJUNEN, S. 1994. Banana allergy in patients with immediate-type hypertsensitivity to natural rubber latex: characterization of cross-reacting antibodies and allergens. J. Allergy Clin. Immunol. **93:** 990–994.
19. M'RAIHI, L., D. CHARPIN, A. PONS, *et al.* 1991. Cross-reactivity between latex and banana. J. Allergy Clin. Immunol. **87:** 129–130.
20. LEHRER, S.B. & M.L. MCCANTS. 1987. Reactivity of IgE antibodies with crustacea and oyster allergens: evidence for common antigenic structures. J. Allergy Clin. Immunol. **80:** 133–139.
21. LEUNG, P.S., W.K. CHOW, S. DUFFEY, *et al.* 1996. IgE reactivity against a cross-reactive allergen in crustacea and mollusca: evidence for tropomyosin as the common allergen. J. Allergy Clin. Immunol. **98(5 Pt. 1):** 954–961.
22. LOPATA, A.L. & C. ZINN & P.C. POTTER. 1997. Characteristics of hypersensitivity reactions and identification of a unique 49-kD IgE-binding protein (Hal-m-1) in abalone (*Haiotis midae*). J. Allergy Clin. Immunol. **100(5):** 642–648.
23. CARRILLO, T., F.R. DE CASTRO, C. BLANCO, *et al.* 1994. Anaphylaxis due to limpet ingestion. Ann. Allergy **73:** 504–508.
24. DAUL, C.B., M. SLATTERY, G. REESE & S.B. LEHRER. 1994. Identification of the major brown shrimp (*Penaeus aztecus*) as the muscle protein tropomyosin. Int. Arch. Allergy Clin. Immunol. **105:** 49–55.
25. SHANTI, K.N., B.M. MARTIN, S. NAGPAL, *et al.* 1993. Identification of tropomyosin as the major shrimp allergen and characterization of its IgE binding epitopes. J. Immunol. **151:** 5354–5363.
26. LEUNG, P.S.C., K.H. CHU, W.K. CHOW, *et al.* 1994. Cloning, expression, and primary structure of *Metapenaeus ensis* tropomyosin, the major heat-stable shrimp allergen. J. Allergy Clin. Immunol. **92:** 837–845.
27. LEUNG, P.S., Y.C. CHEN, D.L. MYKLES, *et al.* 1998. Molecular identification of the lobster muscle protein tropomyosin as a seafood allergen. Mol. Marine Biol. Biotechnol. **7:** 12–20.
28. LEUNG, P.S.C., Y. CHEN, M.R. GERSHWIN, *et al.* 1998. Identification and molecular characterization of *Charybdis feriatus* tropomyosin, the major crab allergen. J. Allergy Clin. Immunol. **102:** 847–852.
29. SANTOS, A.B.R., K.R. TOBIAS, V.P.L. FERRIANI, *et al.* 1999. Identification of tropomyosin from *Periplaneta americana* as a major cockroach allergen. J. Allergy Clin. Immunol. **103(No. 1, Pt. 2):** S122.
30. ASTURIAS, J.A., M.C. ARILLA, N. GOMEZ-BAYON, *et al.* 1999. Molecular characterization of American cockroach tropomyosin (Per a 7), a cross-reactive allergen. J. Immunol. **162:** 4342–4348.
31. AKI, T., T. KODAMA, A. FUJIKAWA, *et al.* 1995. Immunochemical characterization of recombinant and native tropomyosins as a new allergen from the house dust mite, *Dermatophagoides farinae*. J. Allergy Clin. Immunol. **96:** 74–83.
32. REESE, G., R. AYUSO & S.B. LEHRER. 1999. Tropomyosin: an invertebrate pan-allergen? Intl. Arch. Allergy Immunol. **119(4):** 247–258.
33. SMILLIE, L.B. 1979. Structure and functions of tropomyosins from muscle and non-muscle sources. Trends Biochem. Sci. **4:** 151–155.
34. AYUSO, R., S.B. LEHRER & G. REESE. 2002. Identification of continuous, allergenic regions of the major shrimp allergen Pen a 1 (tropomyosin). Int. Arch Allergy Immunol. **127(1):** In press.

35. REESE, G., R. AYUSO, S.M. LEONG-KEE, *et al.* 2001. Epitopes mapping of the major shrimp allergen Pen a 1 (tropomyosin) and mutational substitution analysis. Submitted.
36. AYUSO, R., G. REESE & S.B. LEHRER. 2001. Molecular basis of arthropod cross-reactivity: cross-reactive epitopes of shrimp (Pen a 1), house dust mite (Der p 10) and cockroach (Per a 7) tropomyosins. Submitted.
37. BURKS, A.W. D. SHIN, G. COCKRELL, *et al.* 1997. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. Eur. J. Biochem. **245(s):** 334–339.
38. STANLEY, J.S., N. KING, A.W. BURKS, *et al.* 1997. Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. Arch. Biochem. Biophys. **342(2):** 244–253.
39. RABJOHN, P., E.M. HELM, J.S. STANLEY, *et al.* 1999. Molecular cloning and epitope analysis of the peanut allergen Ara h 3. J. Clin. Invest. **103(4):** 535–542.

511748

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 10:13 AM
To: STIC-ILL
Subject: RE: 09/494,096

Please deliver:

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

Therapeutic Drug Monitoring
15:514–520 © 1993 Raven Press, Ltd., New York

Biotechnology: An Introduction to Recombinant DNA Technology and Product Availability

†‡Ronald P. Evens and *Mark Witcher

**Manufacturing Department and †Professional Services Department, Amgen Inc., Thousand Oaks; and ‡School of Pharmacy, University of Southern California, Los Angeles, California, U.S.A.*

Summary: Exciting advances in biotechnology have led to the development of innovative biological products to improve health care. In this review article, the major techniques of product development and manufacturing in biotechnology are discussed, with a focus on recombinant DNA technology. In addition, monoclonal antibody, nucleotide blockade, polymerase chain reaction, antisense, and gene therapy technologies will be defined briefly. For recombinant-DNA technology, the issues of gene isolation, gene cloning, protein expression, scale-up (manufacturing), and quality assurance are addressed. The 16 approved products and the research pipeline are characterized. In addition, major usage issues for biological products are noted. This review serves as an introduction to the science and applications of biotechnology present and future. **Key Words:** Biotechnology—Recombinant DNA technology—Biological products.

During the past two decades, biotechnology has evolved from a laboratory science to a dynamic research-based business and an established form of therapy that benefits patients everyday. The sophisticated science of biotechnology continues to unfold and is expanding rapidly, from the core technologies of recombinant DNA procedures and monoclonal antibodies to whole new techniques, such as antisense (anti-RNA code blockade) (1). The biotechnology revolution is a success because of the scientific ingenuity and entrepreneurial spirit of molecular biologists, biochemists, and investment capitalists, who have formed >1,000 biotechnology research companies in the United States that develop biological products to mitigate the effects of diseases in new ways.

Two indispensable factors in this success have been substantial research collaboration with universities and research and regulatory cooperation with government agencies. A 1991 survey showed that 742 biotechnology firms had ~$12 billion in revenue (2). In 1992, the top 79 biotechnology companies, according to stock market analysts, generated $2.127 billion in revenue (3), which was corroborated in a recent review article (13 molecules were responsible for $2.14 billion in sales in 1992) (4). The research and development commitment is impressive as well; 742 companies (1991 survey) reportedly invested >$3.63 billion in research (~30% of revenues) (2). From 1990 to the year 2000, the U.S. biotechnology market is projected to grow tenfold, including pharmaceutical, diagnostic, agricultural, and environmental products. This review article focuses on the scientific principles of biotechnology, product development issues, and products currently in use and under development.

DEFINITIONS

The molecules or biological products created by biotechnology research and manufacturing to date

Address correspondence and reprint requests to Dr. Ronald P. Evens at Amgen Inc., 1840 De Havilland Drive, Thousand Oaks, CA 91320, U.S.A.

are all proteins. The manufacturing process for these molecules uses a living system, usually a cell; this cell produces all of its normal proteins plus one additional protein, very similar to naturally occurring human molecule. Van Noordwijk offered a broadly applicable, yet simple definition of biotechnology: "the use of tissue cultures, living cells or cell enzymes to make a defined product" (5). More precise definitions exist and often delineate specific processes that make up biotechnology, such as hybridoma, cell fusion, and recombinant DNA. Furthermore, biotechnology can be described as a composite of several key scientific disciplines—molecular biology, microbiology, biochemistry, immunology, genetics, and engineering.

In the brief time period from 1982 to 1993, biological products have become widely available (15 molecules and 21 products by the fall of 1993). These products are being used in hundreds of thousands of patients annually for such common maladies as diabetes, cancer, heart attacks, and anemia. Biotherapy is a term now used to characterize the clinical use of these products. Its definition encompasses the administration of proteins, which often are duplicates or functional equivalents of naturally occurring molecules in the human body, in order to enhance normal body functions or replace deficient or dysfunctional proteins. Biotherapy is being integrated into patient care along with the established therapies of drugs, diet, surgery, physical procedures, radiation, and psychotherapy.

DESCRIPTION OF RECOMBINANT DNA TECHNOLOGY

Molecular biology is based on the central tenet that DNA makes RNA makes protein (6). This process comprises three major steps: transcription, translation, and completion of the protein. Genes are DNA nucleotide sequences, the basic units of heredity, and the starting material for recombinant DNA (r-DNA) technology. In the 23 pairs of chromosomes in the nucleus, made up of three to four billion base pairs of nucleotides, we need to locate the one gene of 50,000 genes responsible for producing a specific protein. In addition, genes contain coding DNA sequences (exons) responsible for protein production; a leader DNA sequence to initiate transcription; noncoding DNA material (introns), which must be eliminated in the formation of messenger RNA (m-RNA); and nucleotide sequences that indicate the terminal end of a gene and turn off transcription. The end product of r-DNA technology is a protein, which includes a set of amino acids of the correct number, type, and sequence. Several structural requirements are also critical for optimal activity of proteins; these include disulfide cross-linkage, three-dimensional folding, and glycosylation with sugar moieties of the right number, location, and type.

The steps in the process and the basic elements of recombinant DNA technology are outlined in Table 1 and are described herein. Recombinant DNA technology starts with the step of gene isolation, which includes three methods. One method begins with protein analysis; the composition of the protein must be identified fully with regard to the type and sequence of amino acids. Because the DNA codons (triplets) for each amino acid already are known, the DNA sequence then is synthesized in reverse from the protein. A second method involves isolation of the m-RNA that is responsible for producing a specific protein. Then, the viral enzyme, reverse transcriptase, is used to create the complementary DNA for the m-RNA. A third method is a screening process of the full genomic library when the protein cannot be identified fully. From two

TABLE 1. *Steps in process and basic elements of recombinant DNA technology*

Step 1
Gene isolation (three alternatives with different starting material)
From protein analysis to DNA triplets to DNA synthesis
From m-RNA identification plus viral reverse transcriptase to c-DNA formation
From chromosome pool with DNA probes to DNA screening for gene
Step 2
Cloning/expression (from genes to proteins in the laboratory)
Plasmid and gene-starting DNA materials
Restrictive endonuclease enzyme for plasmid opening
DNA ligase enzyme for plasmid closure with human gene
Recombinant plasmid entry into host cell
Host cell/plasmid cloning (duplication)
Gene expression of protein by host cells
Step 3
Scale-up process (manufacturing of commercial product)
Inoculum—daughter cells from master cell bank
Cell culture—serial fermentation—crude protein mixture
Purification—desired protein in bulk
Formulation—final product
Step 4
Quality control measures (tests of products and process)
Plasmid/host cell—integrity assessment
Bulk product—structure, potency, purity, activity
Process validation—efficiency and reproducibility
Final product—for bulk product, plus stability, sterility

separate peptide segments of the protein, two distinct oligonucleotide probes matching the two peptides are created. Next, the chromosome is denatured into several hundreds of thousands of DNA segments. This pool of DNA segments is searched using each oligonucleotide probe. A DNA segment that matches both probes would be a likely clone representing the target gene.

The second step in product development for r-DNA technology is cloning and expression, which is worked out in the laboratory. Cloning is duplication of a gene and its host cells. Expression is production of the protein from the gene in a host cell. The components of this step include the gene, plasmids, restriction endonucleases, DNA ligases, DNA linkers, DNA promoters/enhancers, and host cells.

Plasmids are circular rings of DNA that serve as the vector to carry the human gene into host cells. Key properties of plasmids are that they are DNA material, they have the capacity for autonomous replication of DNA, they allow ready acceptance of genes, they are transferable between cells, and they are stable during the recombinant process. Their source is bacteria, where they are responsible for transference of virulence or resistance, e.g., the pBR322 plasmid. Restriction endonuclease enzymes are employed to cut open plasmids at specific nucleotide triplets; a certain enzyme is chosen to open the plasmid at sites compatible with the terminal triplets of a gene. Bacteria again provide this tool, with ~100–200 distinct enzymes available.

DNA ligase enzyme is used to recombine the DNA material, incorporating the human gene into the bacterial plasmid. If the gene does not permit full closure of the plasmid ring, a DNA linker (segment of DNA) is added to piece together the gene with the plasmid ring. Sometimes a special DNA segment from a viral or bacterial source is added to the gene/plasmid DNA ring to increase gene replication and protein production. This recombined plasmid is then transferred to a host cell.

Host cells need to possess three primary characteristics: a short reproductive cycle, ease of growth in vitro, and acceptability of plasmids. Host cells can be fungal (e.g., yeast), bacterial (e.g., *E. coli*), or mammalian (e.g., Chinese hamster ovary) cells. The host cells with the recombined plasmids are placed in special nutrient mixtures designed for optimal cell viability and protein production. These cells are cloned (duplicated) along with intracellular plasmid multiplication. These host cells retain the capacity to produce all the various proteins in their normal life cycle and express one more protein, which is the duplicate or functional equivalent of the human protein produced by human gene expression. The cloning process results in a master working cell bank of special new parent cells capable of producing human proteins.

The site of manufacture of proteins in r-DNA technology is living cells. Scale-up of protein production is required to manufacture sufficient amounts of protein from very large cell colonies to meet patients' needs in a reasonably cost-effective manner. The manufacturing processes primarily are twofold, based on the type of host cells, that is, mammalian (eucaryote) or microbial (prokaryote) cells. In mammalian cell culture, cells grow slowly, with a doubling time of ~24–36 h. A production lot can require more than a month to produce. Prominent challenges are to maintain an environment that remains sterile, keeps cells viable and productive, and avoids any contamination over a 1–2 month period. Mammalian host cells secrete the proteins extracellularly, which enhances harvesting of proteins from media. Chinese hamster ovary cells are the mammalian cell line employed for epoetin alfa (EPOGEN) and alteplase (Activase). For microbial cell cultures, the cell doubling time can be as little as 30 min, which is highly desirable and, hence, is preferred for producing proteins. Microbial fermentation is an easier process and is more frequently used in r-DNA manufacturing. It is employed for filgrastim (NEUPOGEN). Microbial cells do not secrete proteins extracellularly and store proteins in vesicles, which complicates harvesting of proteins.

Four primary steps are involved in scale-up for mammalian r-DNA manufacturing: inoculum, cell culture, purification, and formulation. The inoculum step entails obtaining a single vial of daughter cells from the parent cells in the master working cell bank. These daughter cells are grown in vitro in small spinner flasks to initiate the manufacturing scale-up. This initial cell culturing expands cell volume with serially larger spinner flasks; cell density in the flasks is the critical factor in maximizing cell growth. The goal of this inoculum phase is to prepare a sufficient number of cells for conduct of the cell culture phase. Cell culture production is the core process in the scale-up in r-DNA technology: the proteins are produced by serial fermentation, in which the amount of cloning (number of functioning cells, volume of growth media, and protein production) are increased substantially in a stepwise fash-

ion. This cell culture process can employ a large fermenter wherein the cells float free in the medium, that is, a suspension culture. Alternatively, attachment technology, e.g., a fluidized bed fermenter or roller bottles, is used, wherein cells are fixed on a surface and are bathed in media.

The fermentation process requires careful control of the cell environment, including the nutrient medium, the air mixture (oxygen, nitrogen, carbon dioxide, and moisture), the breathability of the fermentation flasks (between flasks and ambient environment), the temperature of flasks, and the removal of cellular waste. Media shifts in roller bottles are performed periodically to maintain cell viability. Harvesting of protein from the medium is done at several times during the cell culture process, along with a step to concentrate the liquid mixture to create a crude multi-protein mixture. Throughout this fermentation period, the cells must be kept alive and functional without mutation, change, or contamination (e.g., viruses or oncogenes).

Purification of the crude protein is accomplished with a sequence of chromatographic steps to separate the desired protein from the crude mixture. The physical properties of the proteins (such as the size of molecules, the hydrophobic properties, and electrical charge) allow this separation to be done. The common chromatographic methods include gel-filtration, ion-exchange, affinity, reversed-phase, and hydrophobic techniques. Sterile filtration is then performed, to obtain a purified bulk of target protein without contaminating material or organisms. The final step, formulation, requires mixing the protein with a diluent plus stabilizers and buffers specific for that protein product. Filling and labeling of ready-to-use vials is necessary. Some biological products are lyophilized attempting to improve a product's shelf life, although reconstitution by practitioners then is required before use.

For microbial fermentation, the cell culture step uses one or several large fermenters containing the appropriate medium. Because of the high cell-growth rate, the fermenter has the capacity to force air into the media to maintain cell viability. At the end of fermentation, the cells can be killed without denaturing the proteins that are stored in intracellular inclusion bodies. The contents of the fermenter often are centrifuged to create a cell paste. Recovery of proteins from this microbial system usually calls for extraction of the inclusion bodies from the dead cells. The cells are broken up with a homogenizer or other high-shear device to obtain the inclusion bodies. A high-shear device also is employed to process the inclusion bodies and remove the bulk protein. The next step is purification and formulation similar to what we have described herein.

Quality control is a substantial challenge in the biotechnology industry, with >70 tests employed to ensure the integrity, quality, and activity of the intermediates as well as the final product. Biological products usually are proteins; as such, they are more complex compounds, have a more detailed and sophisticated manufacturing process, and are made in living cells, in contrast to the chemical synthesis and/or extraction processes used for drugs. These three features lead to a potentially greater opportunity for contamination, alteration, or denaturation of the protein product.

Four areas are addressed by quality control, that is, plasmids and host cells, bulk product, process validation, and final product batches. First, a guarantee of the integrity of host cells and plasmids necessitates, for example, chromosomal analysis of karyotype, contamination screens for viruses and oncogenes, and gene stability assessment. Second, testing of bulk product is extensive for a protein encompassing, for example, high-pressure liquid chromatographic analysis, radioimmunoassay, Western and Southern blot analyses, amino acid sequencing, peptide maps, and even bioassay. The goals of such a large number of sophisticated tests are to ensure structure, purity, potency, and activity of the protein. Third, final product is evaluated with the same tests as for bulk products, along with standard assessments of final formulation (such as tests of sterility, excipients, volume, appearance, and stability). Fourth, process validation tests are intended to address the efficiency and reliability of the manufacturing process.

TECHNIQUES OF BIOTECHNOLOGY

Although recombinant DNA technology is responsible for >90% of the available biological products, biotechnology has evolved to create several key technologies for product development, as represented in Table 2. Monoclonal antibody (MAb) technology has been in use for the past 20 years and has led to many diagnostic advances, e.g., ELISA (enzyme-linked immunosorbent assays) test kits for drug testing and OncoScint CR103, radiolabeled monoclonal antibodies for diagnosis of colorectal cancer. Therapeutic applications are limited at this

TABLE 2. *Techniques of biotechnology*

Recombinant DNA technology
Monoclonal antibody technique
Polymerase chain reaction
Gene therapy
Nucleotide blockade
Peptide technology
Carbohydrate technology

time to one product, antirejection (kidney) therapy with murononab, but many products are in development, especially for cancer therapy. Manufacture of a MAb requires creation of a hybridoma from mouse B cells, which produce specific antibodies against a target antigen, and from myeloma cells, which have long lives and produce large amounts of antibodies (7). The hybridoma possesses properties of both cell components. It is cloned and grown to yield antibodies, which are then harvested and purified.

The polymerase chain reaction is a biotechnology whereby substantial amplification (over several 100,000-fold) of a target nucleic acid sequence (e.g., a gene) is obtained (8). This enzymic reaction happens in repeated cycles of a three-step process: (a) DNA is denatured to separate the two strands; (b) a nucleic acid primer is hybridized to each DNA strand at a specific location in the nucleic acid sequence; and (c) DNA polymerase enzyme is added for extension of the primer along the DNA strand to copy the target nucleic acid sequence. Each cycle doubles the DNA molecules copied, and the cycle is repeated until sufficient DNA sequence material is copied, e.g., 20 cycles with a 90% success rate yields a 375,000 amplification of a DNA sequence.

Gene therapy is employed in inheritable diseases where a protein is either dysfunctional or not produced because of an abnormal gene (9). These common and life-threatening diseases include cystic fibrosis, hemophilia, sickle cell anemia, and diabetes. The gene responsible for the malady must be identified first, which was done, for example, for ADA (adenine deaminase) enzyme deficiency. This disease leads to a severely compromised immune deficiency (SCID), often causing death in childhood or adolescence. The normal human gene for the fully functional protein is cloned and inserted into a carrier, such as a harmless virus. A patient's cells, e.g., T lymphocytes, are grown in the laboratory, and the cells receive the gene from the viral carrier. The patient's cells start producing the missing protein to correct the deficiency. These cells with the extra functional gene then are returned to the patient, and the normal protein is produced and released, alleviating the disease. For SCID and other diseases, gene therapy is life-saving.

Nucleotide blockade or antisense focuses on arresting expression of dysfunctional messenger RNA or DNA (10). A complementary m-RNA is created to match an abnormal m-RNA. The two m-RNA strands complex together, preventing translation of the m-RNA to form disease-producing proteins. Anti-DNA strands also can be created to complex with DNA to form a triple helix. Oligonucleotides, or short single strands of nucleic acids, instead of the full m-RNA also can be employed to block RNA expression. Viral disease (herpes simplex and HIV) and cancer (oncogenes) are two targets of this form of biotechnology.

Peptide technology entails screening for polypeptide molecules that can mimic large proteins, thus affording more simple products that may be more stable and easier to produce. The peptides can serve as protein receptor agonists or antagonists. Carbohydrate chemistry is targeted at inflammatory and immune diseases involving leukocyte activation. Expression of carbohydrate molecules from cell surfaces (cell adhesion molecules) can be prevented, arresting cell migration or activation and, in turn, moderating the leukocyte-related components of inflammation.

These brief descriptions of biotechnology techniques now employed in product development expose the vast potential for many and varied biological products, which will modify or correct many diseases. The biological products now in development in clinical trials well exceed 100 molecules.

BIOLOGICAL PRODUCT AVAILABILITY

Fifteen biological molecules have been developed and marketed in the years from 1982 through mid-1993. Fourteen companies are marketing 20 products. Table 3 lists the molecules, products, companies, indications, and year of availability. Some products are used for single indications [such as insulin for diabetes (1982) or Factor VIII for hemophilia (1992)]. However, many products are approved first for one major indication and then, after substantial clinical research into additional indications, are approved for other uses at a later time. Examples include interferon α-2b (Intron A) for hairy cell leukemia (1986), AIDS-related Kaposi's sarcoma (1988), hepatitis C (1991), and hepatitis B (1992), and epoetin α (EPOGEN, Procrit) for ane-

TABLE 3. *Biological products*[a]

Generic name	Trade name	Therapeutic area
Human insulin	Humulin (Lilly)	Insulin-dependent diabetes mellitus (1982)
Human growth hormone	Protropin (Genentech) Humatrope (Lilly)	Human growth hormone deficiency in children (1985)
Hepatitis B vaccine	Engerix-B (SmithKline Beecham) Recombivax HB (MSD)	Hepatitis B prophylaxis (1986)
Interferon α-2a	Roferon A (Hoffman LaRoche)	Hairy cell leukemia (1986); AIDS-related Kaposi's sarcoma (1988)
Interferon α-2b	Intron A (Schering-Plough)	Hairy cell leukemia (1986); AIDS-related Kaposi's sarcoma (1988); chronic hepatitis types B (1992) and C (non-A, non-B) (1991); condylomata acuminata (1988)
Muromonab-CD3	Orthoclone OKT 3 (Ortho Biotech)	Acute allograft rejection in renal transplant patients (1986)
Alteplase	Activase (Genentech)	Acute myocardial infarction (1987); pulmonary embolism (1990)
Epoetin α	Epogen (Amgen) Procrit (Ortho Biotech)	Certain anemias—chronic renal disease (1989); AIDS (1991); cancer chemotherapy (1993)
Interferon α-n3	Alferon N (Interferon Sciences)	Condylomata acuminata (1989)
Interferon γ-1b	Actimmune (Genentech)	Chronic granulomatous disease (1990)
Filgrastim (G-CSF)	Neupogen (Amgen)	Febrile neutropenia (infection) due to myelosuppressive chemotherapy (1991)
Sargramostim (GM-CSF)	Leukine (Immunex) Prokine (Heochst-Roussel)	Myeloid reconstitution after bone marrow transplantation (1991)
Aldesleukin	Proleukin (Chiron)	Metastatic renal cell carcinoma (1992)
Factor VIII	KoGenate (Miles) Recombinate (Baxter)	Hemophilia A (1992)
Interferon β	Betaseron (Chiron, Berlex)	Multiple sclerosis (1993)

G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte–macrophage colony-stimulating factor.

[a] List of products is given in order of date of first product approval.

mia stemming from chronic renal disease (1989), from AIDS-related antiviral therapy (1991), and from cancer chemotherapy (1993). The disease entities treated by these 15 molecules are commonplace and address most organ systems and 18 separate conditions. They can be used in endocrinology (diabetes, growth), hepatitis (types B and C), oncology (hairy cell leukemia, AIDS-related Kaposi's sarcoma, renal cell carcinoma, neutropenia, and infections from chemotherapy), anemias, cardiology (heart attacks, pulmonary embolism, hemophilia), neurology (multiple sclerosis), and organ (kidney) rejection. This brief description establishes the profound impact of biotechnology on health care, bringing therapeutic agents to treat diseases that often were previously untreatable. A new arm of therapy, biotherapy, has been created.

FUTURE BIOTECHNOLOGY PRODUCTS

The first decade of biotechnology drug development produced an admirable record of one new biological molecule per year, on average. This early success is expected to be followed by a host of products in similar and new therapeutic areas. According to 1991 Pharmaceutical Manufacturing Association (PMA) data, 132 biological products were in clinical development and 21 molecules were at the Federal Drug Administration (FDA) for regulatory review and potential approval (11). The time frame to bring such new biological products to market (patient availability) has been documented to average ~4.8 years in the United States and 3 years in Europe; two-thirds of this time is consumed by clinical research and one-third by regulatory review

(12). According to PMA data, from 1982 to 1992, FDA review of product license applications for biological products averaged 24 months, and new indications took 30 months for review and approval. Over the past 5 years, the research pipeline has grown from <80 to >130 products (60% growth), and the expected time required for product research and approval is ~5 years.

These new biotherapies incorporate both extensions of current product categories and whole new entities. Table 4 cites 10 product categories for future biological products and includes >150 new products or indications that are in development (13). Most therapeutic areas are being targeted by biotechnology, but oncology has the largest number of products; some two-thirds to three-quarters of interferons, interleukins, and monoclonal antibodies are used in this area. In vaccines, AIDS is the major focus. Tissue-growth factors, including epidermal growth factor, fibroblast growth factor, and platelet-derived growth factor, address repair of wounds due to injury, surgery, chronic disease, or decubiti. Clotting factors include proteins to improve clotting and to dissolve clots. The last category (labeled "various") includes such molecules as dismutases, tumor necrosis factor, and soluble CD4.

TABLE 4. *Future biotherapies*

Category	Number of indications and products
Antisense	5
Clotting factors	8
Hematopoietic factors	12
Hormones	7
Interferons	16
Interleukins	13
Monoclonal antibodies	58
Tissue growth factors	11
Vaccines	18
Various	>15

See Burrill and Lee for summary (13).

Biological products have become the mainstay of therapy for major diseases already benefiting millions of patients each year, with only 15 molecules and 20 products in use to date. Future products and indications exceed 100 in number. The full impact of these therapies and the benefits to patients still are being discovered. Biotherapy holds the promise of a natural therapy, wherein proteins, nucleotides, and other human molecules are recreated and employed to ameliorate or even cure disease. We are on the threshold of a pharmacologic and therapeutic revolution.

REFERENCES

1. Drews J. Into the 21st century: biotechnology and the pharmaceutical industry in the next ten years. *Bio/technology* 1993;11:S16–S20.
2. Dibner MD. Tracking trends in the U.S. biotechnology. *Biol technology* 1991;9:1334–7.
3. Miller L, Escrow J, Malloy M, eds. *Biotechnology industry evaluation: 1992 sector analysis*. New York: Paine Webber Inc., Health Care Group, 1992.
4. Klausner A. Back to the future: biotech product sales 1983–1993. *Bio/technology* 1993;11:S35–6.
5. Van Noordwijk J. Quality assurance of products manufactured by recombinant DNA technology: introduction and elements of a philosophy. *Arzneimittalforschung* 1988;38: 943–7.
6. Sattele DB, ed. *Biotechnology in perspective*. London: Hobsons Publishing, 1990.
7. Pirofski L, Gasadevall A, Rodriguez L, Zuckier LS, Scharff MD. Current state of hybridoma technology. *J Clin Immunol* 1990;10(suppl. 6):5s–14s.
8. Lipner M. The polymerase chain reaction: amplifying its role in research and beyond. *Oncol Times* 1992;14–5.
9. Goldspiel BR, Green L, Calis KA. Human gene therapy. *Clin Pharmacy* 1993;12:488–505.
10. Nagel KM, Holstad SG, Isenberg KE. Oligonucleotide pharmacotherapy: an antigene strategy. *Pharmacotherapy* 1993; 13:177–88.
11. In development: biotechnology medicines. Washington, D.C.: Pharmaceutical Manufacturing Association, 1991.
12. Bienz-Tadmor B. Biopharmaceuticals go to market: patterns of worldwide development. *Bio/technology* 1993;11:168–72.
13. Burrill GS, Lee Jr KB. *Biotech '92: promise to reality*. An industry annual report. San Francisco: Ernst and Young, 1991.

STIC-ILL

5/1750

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 10:13 AM
To: STIC-ILL
Subject: RE: 09/494,096

Please deliver:

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

dimensional structure of immunoglobulins. *Annu. Rev. Biochem.*, 44, 639-667.

Deisenhofer, J. (1981), Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of a protein A from *Staphylococcus aureus* at 2.9 Å and 2.8 Å resolution. *Biochemistry*, 20, 2361-2370.

Guddat, L.W., Herron, J.N. & Edmundson, A.B. (1993), Three-dimensional structure of a human immunoglobulin with a hinge deletion. *Proc. Nat. Acad. Sci.* (Wash.), 90, 4271-4275.

Harris, L.J., Larson, S.B., Hasel, K.W., Day, J., Greenwood, A. & McPherson, A. (1993), The three-dimensional structure of an intact monoclonal antibody for canine lymphoma. *Nature* (Lond.), 360, 369-372.

Köhler, G. & Milstein, C. (1975), Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* (Lond.), 256, 495-497.

Marquart, M., Deisenhofer, J. & Huber, R. (1980), Crystallographic refinement and atomic models of the intact immunoglobulin Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. *J. Mol. Biol.*, 141, 369-391.

Padlan, E.A., Davies, D.R., Rudikoff, S. & Potter, M. (1976), Structural basis for the specificity of phosphorylcholine-binding immunoglobulins. *Immunochemistry*, 13, 945-949.

Poljak, R.J., Amzel, L.M., Avey, H.P., Chen, B.L., Phizackerley, R.P. & Saul, F. (1973), Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8 Å resolution. *Proc. Nat. Acad. Sci.* (Wash.), 70, 3305-3310.

Segal, D.M., Padlan, E.A., Cohen, G.H., Rudikoff, S., Potter, M. & Davies, D.R. (1974), The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proc. Nat. Acad. Sci.* (Wash.), 71, 4298-4302.

Schiffer, M., Girling, R.L., Ely, K.R. & Edmundson, A. (1973), Structure of a λ-type Bence-Jones Protein at 3.5 Å resolution. *Biochemistry*, 12, 4620-4631.

Sarma, V.R., Silverton, E.W., Davies, D.R. & Terry, W.D. (1971), The three-dimensional structure at 6.0 Å resolution of a human γG1 immunoglobulin molecule. *J. Biol. Chem.*, 246, 3753-3759.

Wu, T.T. & Kabat, E.A. (1970), An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.*, 132, 211-250.

Zhao-chang Fan, Lin Shan, Guddat, L.W., Xiao-min He, Gray, W.R., Raison, R.L. & Edmundson, A.B. (1992), Three-dimensional structure of an Fv from a human IgM immunoglobulin. *J. Mol. Biol.*, 228, 188-207.

Effects of amino acid sequence changes on antibody-antigen interactions

P.M. Colman

Biomolecular Research Institute, 343 Royal Parade, Parkville, 3052 (Australia)

Amino acid substitutions in antibody-antigen interfaces play an important role in affinity maturation of antibody responses and in antigenic variation. Structural studies show that some non-conservative changes are tolerated in these interfaces. By analogy with observations of amino acid exchangeability within homologous or mutated protein structures, this suggests that the number of different antibody specificities is less than the total number of antibodies which can be encoded by the various genetic mechanisms. On the other hand, conformational adaptability of antibody structures during binding of antigen suggests that one antibody can serve to bind a number of different antigens.

Single amino acid sequence changes within the interface of an antibody-antigen complex are important in two biological contexts. Firstly, such changes in the antibody have the capacity to drive the affinity towards more tightly bound complexes. Secondly, such changes in the antigen can effectively abolish the interaction entirely, providing an effective mechanism for antigenic variation. Although antibody and antigen display biologically asymmetric behaviour in regard to the effects of single amino acid substitutions, in principle and in practice, substitutions in either partner can raise or lower the affinity. Tolerance of amino acid sequence substitutions within antibody-antigen interfaces suggests that they

be considered examples either of protein-protein interactions generally or of the core of folded protein structures. In those cases, degeneracy of amino acid sequence information is well documented among families of homologous proteins, and it suggests that the number of different antibody specificities is less than the estimated number of different antibody molecules.

Antigenic variation

Substitutions which result in loss of binding are not amenable to direct study of complexes, although several studies have sought to rationalize the effects of the new amino acid in the context of the complex involving the "parent" amino acid.

Escape mutants of viral antigens, selected by growth of virus in the presence of monoclonal antibody, provide many examples of the type of substitution which can render the antigen unrecognizable by the selecting antibody (see, for example, Webster *et al.*, 1987). There are no discernible trends in these data which suggest preference for replacement of residues in one physico-chemical class by those in another as a preferred means of abolishing binding (Colman, 1992). Studies of the three-dimensional structures of a number of these antibody selected mutants shows that local changes, at and immediately adjacent to the site of the mutated residue, suffice to disrupt the interaction with the selecting antibody (Knossow *et al.*, 1984; Varghese *et al.*, 1988; Tulip *et al.*, 1991).

Tulip (1990) has observed that a variety of escape mutants of the influenza virus neuraminidase, selected by the anti-neuraminidase antibody NC41, result in shape changes of the antigen which could interfere with surface docking if the remainder of the antigen and the antibody were rigid and unable to relax around the mutated residue. In another case, a hydrogen bond is potentially lost by substitution to a smaller residue (Tulip, 1990). Bhat *et al.* (1990) report that substitution of Gln by His in lysozyme also causes shape changes which cannot be accommodated by the D1.3 anti-lysozyme antibody, and that, in addition, a hydrogen bond would be lost from the interaction. In both of these examples, the key to the failure of the mutant to bind may reside as much in the inflexibility of the surrounding structures as it does in the physico-chemical environment within the interface of the mutated residue.

Structural data on the effects of point changes in the binding site are available in only one system at present, the influenza virus neuraminidase (Tulip *et al.*, 1992). Two mutants of the N9 subtype neuraminidase, selected with antibodies other than NC41, involve substitutions within the binding site on N9 for NC41. In each case, the affinity is reduced by one to two orders of magnitude (Webster *et al.*, 1987; Gruen *et al.*, 1993). One of these mutants, Asn 329 to Asp, is located near the edge of the binding site for the NC41 antibody and is accommodated by re-positioning the side chain of residue 329 towards the solvent exposed perimeter of the antibody-antigen interface. The other, Ile 368 to Arg, results in a complex pattern of concerted movements around the mutation site to accommodate the arginyl residue. These changes include a shift by 3Å in the position of the arginine from its location in the uncomplexed neuraminidase structure and a shift by more than 1Å of a histidine on the antibody. In both of these structures, the effects of the mutations within the interface are reminiscent of the effects of amino acid substitutions within the interior of protein molecules (Matthews, 1991), Anderson *et al.* (1993) where localized structural rearrangements around the mutation site are frequently observed. Analysis of structures in the Brookhaven data base (Bernstein *et al.*, 1977) suggests a somewhat looser packing density of atoms within an antibody-antigen interface compared to other protein-protein interfaces or the interior of protein molecules (Tulip *et al.*, 1992; Lawrence and Colman, 1993). This suggests, in turn, that antibody-antigen interfaces should not be more sensitive to amino acid substitutions than other protein-protein interfaces.

Somatic mutation

Point mutations accumulating within the variable domains of antibody heavy and light chains are associated with increasing affinity of the antibody for antigen.

Alzari *et al.* (1990) have demonstrated that when 2-phenyloxazolone binds to an antibody, it makes direct contact with amino acids of the V_L domain which are known to be frequently mutated in antibodies with increased affinity for the antigen. In contrast, in this same system, somatic mutation in the V_H domain in no case maps to amino acids directly involved in binding the hapten.

Engineered substitutions in an anti-*p*-azophenylarsonate antibody (Sharon, 1990) have demonstrated that three of nineteen somatic changes observed in the V_H domain suffice to increase the affinity by a factor of 200. Comparison with the three-dimensional structure of the antibody (Rose *et al.*, 1990) suggests that, in every case, the effect must be indirect, since none of the three substituted residues is likely to be in direct contact with the antigen. In this example, there is not yet a report of the antibody-antigen complex structure, nor is the antibody structure highly refined, so the conclusion should be considered preliminary.

Structural context and repertoire size

The above examples paint a confusing picture of the specificity of antibody-antigen interaction. In one structural context, a very conservative substitution may abolish binding; in another, a non-conservative substitution may have very little effect on the binding affinity. How should these observations be treated if one is to estimate the effective size of the antibody repertoire from genetic sources alone? One approach is to treat the antibody-antigen interface the same way as the interior of a protein structure, and to bring to bear on this problem current knowledge about the degeneracy of "structural information" among the twenty amino acids.

Current estimates of the potential number of antibody molecules that can be generated by all the known genetic mechanisms is in excess of 10^{18} (Hunkapiller and Hood, 1989). This and similar other estimates assume each of the 20 amino acids is different from every other amino acid, which is appropriate for purposes of enumeration but not for the purpose of estimating how many different antibody specificities can be produced by an animal.

Data from exchangeability matrices (Dayhoff *et al.*, 1978) or from degeneracy of information in amino acid sequences resulting from overlapping genes (Sander and Schulz, 1979) suggest that, for structural purposes, or more precisely for folding purposes, there are effectively only four or five "different" amino acids. Can the same reduction be applied to binding interactions between proteins? Folding could be considered to be more tolerant of amino acid substitutions because of the cooperative effects caused by the interactants being covalently linked to each other. Binding interactions could be considered less tolerant because the changes involved occur in what might be called the active site. Outside of the antibody-antigen system, mapping of protein-protein interacting surfaces by mutational analysis is also generally successful (Bowie *et al.*, 1990), but some unexpected and unexplained findings do occur (de Voss *et al.*, 1992), suggesting a measure of tolerance of amino acid substitutions generally in protein-protein interfaces.

These arguments affect considerations of antibody repertoire sizes. In the facile extreme of substituting five (types of amino acids) for twenty in the calculations of Hunkapiller and Hood (1989), the expected number of different antibody specificities is reduced to order 10^8. Another way to estimate the number of different specificities is to argue from the physical size of a typical binding site on antibody for antigen. For protein antigens, this surface size is of the order of 15 amino acids. The numbers of different specificities that can be encoded over such a surface based on twenty or five structurally different amino acids are order 10^{19} and 10^{10}, respectively, if one ignores the important influences of overall surface shape and partiality of solvent exposure of amino acids within the surface. Similar arguments and estimates of repertoire size are applicable to T-cell receptors.

Some compromise in the capacity of the immune system to cope with "foreign" structures is implied by these degeneracy arguments. The above estimates have carried the argument to the extreme and could therefore be viewed as an extreme lower limit of repertoire size. Nevertheless, genetic sources of diversity are only part of the story. Antibodies, as proteins, display the usual types of conformational adaptability in binding to ligands, as do other proteins, *i.e.* side chain rearrangements and main chain changes within loop structures (Colman, 1988; Wilson and Stanfield, 1993). In addition, there is growing experimental evidence for the functioning of the V_L-V_H interface as a structural adaptor allowing movements of the heavy chain CDR *en masse* with respect to the light chain CDR during engagement with antigen (Colman, 1988, 1991; Herron *et al.*, 1991; Bhat *et al.*, 1990; Stanfield *et al.*, 1993). These changes in antibody structure are believed to be specific to the interacting antigen. Different antigens may therefore induce different structural responses in the same antibody, adding a structural dimension to diversity (Colman, 1988).

References

Alzari, P.M., Spinelli, S., Mariuzza, R.A., Boulot, G., Poljak, R.J., Jarvis, J.M. & Milstein, C. (1990), Three-dimensional structure determination of an anti-2-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response. *EMBO J.*, 9, 3807-3814.

Anderson, D.E., Hurley, J.H., Nicholson, H., Baase, W.A. & Matthews, B.W. (1993), Hydrophobic core repacking and aromatic-aromatic interaction in the thermostable mutant of T4 lysozyme Ser117→Phe. *Prot. Sci.*, 2, 1285-1290.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Schimanouchi, T. & Tasumi, M. (1977), The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112, 535-542.

Bhat, T.N., Bentley, G.A., Fischmann, T.O., Boulot, G. & Poljak, R.J. (1990), Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. *Nature* (Lond.), 347, 483-485.

Bowie, J.U., Reidhaar-Olson, J.F., Lim, W.A. & Sauer, R.T. (1990), Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science*, 247, 1306-1310.

Colman, P.M. (1988), Structure of antibody-antigen complexes: implications for immune recognition. *Adv. Immunol.*, 43, 99-132.

Colman, P.M. (1991), Antigen-antigen receptor interactions. *Curr. Opin. Struct. Biol.*, 1, 232-236.

Colman, P.M. (1992), Structural basis of antigenic varia-

tion: studies of influenza virus neuraminidase. *Immunol. Cell Biol.*, 70, 209-214.
Dayhoff, M.O., Schwartz, R.M. & Orcutt, B.C. (1978), Atlas of protein sequence and structure, Vol. 5, Suppl. 3, pp. 345-362. National Biochemical Foundation, Georgetown University Medical Center, Washington, DC.
De Vos, A., Ultsch, M. & Kossiakoff, A.A. (1992), Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science*, 255, 306-312.
Gruen, L.C., McKimm-Breschkin, J.L., Caldwell, J.B. & Nice, E.C. (1993), Affinity ranking of influenza neuraminidase mutants with monoclonal antibodies using an optical biosensor: comparison with ELISA and slot blot assays. *J. Immunol. Methods* (in press).
Herron, J.N., He, X.M., Ballard, D.W., Blier, P.R., Pace, P.E., Bothwell, A.L.M., Voss, W.E. Jr & Edmundson, A.B. (1991), An autoantibody to single-stranded DNA: comparison of the three-dimensional structures of the unliganded Fab and a deoxynucleotide-Fab complex. *Prot. Struct. Func. Genet.*, 11, 159-175.
Hunkapiller, T. & Hood, L. (1989), Diversity of the Immunoglobulin Gene Superfamily. *Adv. Immunol.*, 44, 1-61.
Knossow, M., Daniels, R.S., Douglas, A.R., Skehel, J.J. & Wiley, D.C. (1984), Three-dimensional structure of an antigenic mutant of the influenza virus haemagglutinin. *Nature* (Lond.), 311, 678-680.
Lawrence, M.J. & Colman, P.M. (1993), Shape complementarity at protein-protein interfaces. *J. Mol. Biol.* (in press).
Matthews, B.W. (1991), Mutational analysis of protein stability. *Curr. Opin. Struct. Biol.*, 1, 17-21.
Rose, D.R., Strong, R.K., Margolies, M.N., Gefter, M.L. & Petsko, G.A. (1990), Crystal structure of the antigen-binding fragment of the murine anti-arsonate monoclonal antibody 36-71 at 2.9-A resolution. *Proc. Natl. Acad. Sci.* (Wash.), 87, 338-342.
Sander, C. & Schulz, G.E. (1979), Degeneracy of the information contained in amino acid sequences: evidence from overlaid genes. *J. Mol. Evol.*, 13, 245-252.
Sharon, J. (1990), Structural correlates of high antibody affinity: three engineered amino acid substitutions can increase the affinity of an anti-*p*-azophenylarsonate antibody 200-fold. *Proc. Natl. Acad. Sci.* (Wash.), 87, 4814-4817.
Stanfield, R.L., Takimoto-Kamimura, M., Rini, J.M., Profy, A.T. & Wilson, I.A. (1993), Major antigen-induced domain rearrangements in an antibody. *Structure*, 1, 83-93.
Tulip, W.R. (1990), Crystallographic refinement of a neuraminidase-antibody complex. Ph.D. thesis, University of Melbourne, Australia.
Tulip, W.R., Varghese, J.N., Baker, A.T., van Donkelaar, A., Laver, W.G., Webster, R.G. & Colman, P.M. (1991), Refined atomic structures of N9 subtype influenza virus neuraminidase and escape mutants. *J. Mol. Biol.*, 221, 487-497.
Tulip, W.R., Varghese, J.N., Webster, R.G., Laver, W.G. & Colman, P.M. (1992), Crystal structures of two mutant neuraminidase-antibody complexes with amino acid substitutions in the interface (1992). *J. Mol. Biol.*, 227, 149-159.
Varghese, J.N., Webster, R.G., Laver, W.G. & Colman, P.M. (1988), Structure of an escape mutant of glycoprotein N2 neuraminidase of influenza virus A/Tokyo/3/67 at 3 Å. *J. Mol. Biol.*, 200, 201-203.
Webster, R.G., Air, G.M., Metzger, D.W., Colman, P.M., Varghese, J.N., Baker, A.T. & Laver, W.G. (1987), Antigenic structure and variation in an influenza virus N9 neuraminidase. *J. Virol.*, 61, 2910-2916.
Wilson, I.A. & Stanfield, R.L. (1993), Antibody-antigen interactions. *Curr. Opin. Struct. Biol.*, 3, 113-118.

Recognition of carbohydrates by antibodies

M. Cygler

Biotechnology Research Institute, National Research Council, 6100 Royalmount, Montréal, Québec H4P 2R2 (Canada)

The recognition by antibodies of complex polysaccharides forming the outer layer of cell walls is an integral part of the immune response to bacterial infection and invasion by non-self cells. For these reasons, the interactions of antibodies with sugars have been extensively investigated by physicochemical and immunochemical methods (Young *et al.*, 1983; Bundle, 1989; Glaudemans, 1991; Sigurskjold and Bundle, 1992). The conformation of free (Bush, 1992) and antibody-bound oligosaccharides has been studied by NMR spectroscopy (Glaudemans, 1991; Bundle *et al.*, submitted) and crystallography (Cygler *et al.*, 1991; Vyas *et al.*, 1993). NMR measurements showed that in some cases, binding to an antibody

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al,* J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al,* Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al,* Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al,* Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

Antagonistic peptides specifically inhibit proliferation, cytokine production, CD40L expression, and help for IgE synthesis by Der p 1–specific human T-cell clones

Stephan Fasler, PhD, Gregorio Aversa, PhD, Jan E. de Vries, PhD, and Hans Yssel, PhD *Palo Alto, Calif.*

Background: Allergic disorders are characterized by IgE antibody responses to a multitude of allergens as a result of the ability of these antibodies to specifically bind to high-affinity IgE receptors on mast cells and basophils. This interaction results in receptor activation and release of soluble mediators such as histamine and leukotrienes, which cause allergic reactions in various target organs. Because the synthesis of IgE is tightly regulated by cytokines and CD40 ligand (L) interactions, $CD4^+$ helper T cells are obvious targets, with the aim to modulate allergen-induced IgE responses.

Objectives: Because of the central role of allergen-specific T-helper type 2 (T_{H2}) cells in the pathway leading to IgE synthesis in vitro and in vivo, we have evaluated the possibility of inhibiting allergen-induced activation of these cells by using allergen-derived peptides that have been modified by single amino acid substitutions.

Methods: Three cloned human T_{H2}-like $CD4^+$ T-cell lines, specific for Der p 1, the major allergen in house dust, were used in this study. Upon activation with Der p 1 or specific Der p 1–dervived wild-type peptides, these T-cell clones produce high levels of IL-4 and IL-5 and low levels of interferon-γ and IL-2, respectively, and furthermore give help to B cells for the production of IgE in vitro. Modified synthetic peptides were generated by the introduction of single amino acid substitutions into two different T-cell activation–inducing epitopes on Der p 1. The effects of these modified peptides were studied in Der p 1–induced proliferation, cytokine production, and in vitro IgE production assays.

Results: Several substituted Der p 1–derived peptides failed to induce T-cell proliferation, in contrast to the native peptides. In addition, some of these peptides acted as antagonists by strongly inhibiting wild-type peptide-induced proliferation as well as the production of interferon-γ, IL-2, IL-4, and IL-5, although the production of the latter two cytokines was less affected than that of interferon-γ, even at a 100-fold molar excess of the antagonistic peptides. In addition, the presence of an excess of each of the antagonistic peptides during the activation of Der p I–specific T-cell clones prevented induction of CD40L expression, resulting in a failure of these cells to give help to B cells for the production of IgE in vitro, even in the presence of exogenous IL-4.

Conclusions: Substitution of certain amino acid residues in immunogenic Der p 1–derived peptides results in the generation of peptides that fail to induce proliferation of Der p 1–specific T-cell clones. In addition, these modified peptides have strong antagonistic activities on Der p 1–induced proliferation, cytokine production, and CD40L expression by allergen-specific T-cell clones as well as on T cell–mediated IgE production by B cells. These findings suggest that modified peptides interfere with allergen-induced activation of T cells, including the production of cytokines and the expression of surface molecules important for successful T cell–B cell interactions, and may therefore have therapeutic potential by inhibiting the expansion and function of allergen-specific T_{H2} cells. (J Allergy Clin Immunol 1998;101:521-30.)

Key words: Der p 1, allergen-specific T lymphocytes, allergen-derived peptides, cytokine production, CD40L, IgE synthesis

Allergic disorders are characterized by the presence of antibodies of the IgE isotype, which in house dust mite–induced atopy are mainly directed to the *Dermatophagoides pteronyssinus* group 1 (Der p 1) and group 2 (Der p 2) allergens.[1-3] The synthesis of IgE by B cells in vitro is specifically induced by IL-4 and IL-13,[4-6] together with a T cell–B cell contact-mediated signal, accounted for by the interaction of CD40 and CD40 ligand (L).[7-9] The pivotal role of CD40L in induction of IgE isotype switching and IgE synthesis has been demonstrated in patients with hyper-IgM syndrome who have mutations in their genes encoding CD40L and, as a result, have no or minimal levels of IgG, IgA, and IgE in their circulation.[10-13] Although CD40L is expressed on all activated T cells, $CD4^+$ T-helper type 2 (T_{H2}) cells are considered to play a pivotal role in house dust mite–induced atopy because of their capacity to produce high levels of IL-4 and IL-13.[14-16] Therefore Der p1–specific T_{H2} cells are obvious targets for immune intervention in house dust mite–related atopic diseases and prevention of activation, and subsequent cytokine production by these cells may be beneficial to the clinical status of atopic patients.

Allergen-specific $CD4^+$ T cells recognize peptide antigens, presented by antigen-presenting cells (APCs) in the context of a major histocompatibility complex (MHC) class II molecule. On the basis of crystallographic studies, it has been demonstrated that all but one of the amino acid side-chains within the peptide-binding groove of the MHC class II protein can interact with the T-cell receptor (TCR) upon contact between

From the Human Immunology Department, DNAX Research Institute for Molecular and Cellular Biology, Palo Alto.

Received for publication Jan. 29, 1997; revised Dec. 1, 1997; accepted for publication Dec. 3, 1997.

Reprint requests: Hans Yssel, PhD, INSERM U454, Hôpital Arnaud de Villeneuve, 371 Avenue Doyen Gaston Giraud, 34295 Montpellier CEDEX 5, France.

0091-6749/98 $5.00 + 0 1/1/87930

Abbreviations used

APC:	Antigen-presenting cell
CD40L:	CD40 ligand
EBV-LCL:	Epstein-Barr virus–transformed lymphoblastoid B-cell lines
IFN:	Interferon
mAb:	Monoclonal antibody
MHC:	Major histocompatibility complex
TCR:	T-cell receptor
T_{H2}:	T-helper type 2 (cells)

the $CD4^+$ T cell and the APC.[17, 18] Depending on the positions of replaced amino acids, however, the change will affect the interaction of the peptide with the peptide-binding grove of the MHC class II protein, the interaction of the peptide–class II complex with the TCR, or both.

In addition, there is increasing evidence that substitution of certain amino acids of immunogenic peptides will induce qualitative changes in T-cell responses as a result of altered signaling by the TCR upon its interaction with the peptide–MHC complex. Stimulation of mouse T cells with antigenic peptides containing single amino acid substitutions reportedly resulted in TCR antagonism, segregation of proliferative and cytokine responses, or even induction of anergy.[19-24] Also, in human systems, stimulation of T-cell clones with altered peptides has been shown to result in changes in T-cell responses. Peptides with a single amino acid substitution, which acted as TCR antagonists, were found to selectively induce the production of transforming growth factor-β1 by myelin basic protein–specific human T-cell clones but were no longer able to induce the production of IL-2, IL-4, IL-10, and interferon (IFN)-γ by these cells.[25] Recently, it was reported that stimulation of T-cell clones with altered peptides, derived from dominant epitopes on the house dust mite allergen Der p 2 allergen,[26] or the Japanese cedar pollen allergen Cry j 1,[27] respectively, resulted in the selective enhancement of IFN-γ but not of IL-4 production or induction of cytokine production in the absence of proliferation.[26] The above-mentioned studies show that single amino acid substitutions in T-cell activation–inducing epitopes on major allergens may modulate the cytokine production profile of allergen-specific T cells. However, no information has been provided on the capacity of these peptides to inhibit in vitro IgE synthesis. Moreover, the effects of modified peptides on T cell–B cell–mediated interactions, which are required for productive IgE synthesis, have not been reported yet.

Therefore in this study we have investigated the capacity of peptides, derived from two minimal T-cell activation–inducing epitopes on the major house dust mite allergen Der p 1 and modified by single amino acid substitutions to block the activation of human Der p 1–specific T-cell clones and to inhibit the production of T_{H2} cytokines, the expression of CD40L and T cell–mediated IgE synthesis by B cells in vitro.

METHODS

Cells

The human $CD4^+$ T_{H2}-like, HLA-DR1101–restricted, T-cell clones NP-12 and NP-44, specific for the p94-104 sequence of Der p 1, have been described previously.[15] The $CD4^+$ T_{H2} clone AR-142 recognizes the Der p 1–derived peptide p171-182 in the context of HLA-DR1501 (S. Fasler, unpublished results). All Der p 1–specific T-cell clones were generated by stimulation of peripheral blood mononuclear cells, obtained from atopic patients who were allergic to Der p 1, with 1 μg/ml of a lyophilized extract of Der p 1 (ALK, Bornholm, Denmark), followed by cloning by using a limiting dilution method, as has been described.[15] The T-cell clones were cultured at a concentration of 2×10^5 cells/ml in 24-well Linbro plates (Flow, McLean, Va.) after stimulation with a feeder cell mixture consisting of 10^6 cells/ml irradiated (40 Gray) allogeneic peripheral blood mononuclear cells, 10^5 cells/ml of the irradiated (50 Gray) Epstein-Barr virus–transformed lymphoblastoid B cell line (EBV-LCL) JY, and 0.1 μg/ml PHA (Wellcome Diagnostics, Dartford, U.K.), as described previously.[28] Three to 4 days after each restimulation with feeder cells, the cultures were expanded in culture medium containing 20 U/ml rIL-2 and 200 U/ml rIL-4 (kindly provided by Dr. G. Zurawski, DNAX Research Institute). All cloned T-cell lines and EBV-LCL were free of mycoplasma and were cultured in Yssel's medium, supplemented with 1% human serum.[29] Experiments were carried out in the same culture medium unless indicated otherwise. HLA DR1101–matched monocytes to be used as APC for NP-44 cells (generously provided by Drs. C. Figdor and R. Huybens, Academisch Ziekenhuis Nijmegen, The Netherlands) were isolated from the peripheral blood of a healthy donor by centrifugal elutriation as described previously.[30]

Proliferation assay

Cellular proliferation was measured with an [^{3}H]TdR assay as has been previously described.[15] In brief, 10^5 T cells were mixed with 5×10^4 irradiated (40 Gy) HLA-matched monocytes or 5×10^4 irradiated (50 Gy) autologous EBV-LCL that had been pulsed for 3 hours with specific peptide at concentrations ranging from 0.1 to 10 μg/ml in a flat-bottomed, 96-well tissue culture plate (Flow Laboratories) in a total volume of 200 μl. After 72 hours of incubation at 37° C, 37 kBq [^{3}H]TdR (Amersham, Arlington Heights, Ill.) was added to each well. Six hours later, the cells were harvested onto glass fiber strips and the amount of incorporated [^{3}H]TdR was measured by liquid scintillation counting. The results are expressed as the mean of triplicate cultures ± the standard deviation. Inhibition of proliferation by substituted peptides was calculated by the following formula:

$$\frac{\text{Proliferation with wild type peptide} - \text{spontaneous proliferation}}{\text{proliferation with antagonistic and wild type peptide} - \text{spontaneous proliferation}} \times 100\%$$

Stimulation of T-cell clones and assays for cytokine production

One million T-cell clones per milliliter were cocultured with 2×10^6 autologous EBV-LCL that had been pulsed for 3 hours with specific peptide (0.1 to 10 μg/ml) in 24-well Linbro tissue culture plates. After 48 hours of activation, supernatants were

TABLE I. Substituted peptides derived from two native Der p 1 sequences

Wild-type	A	V	N	I	V	G	Y	S	N	A	Q	G
	1	1	1	1	1	1	1	1	1	1	1	1
171-182	7	7	7	7	7	7	7	7	7	8	8	8
	1	2	3	4	5	6	7	8	9	0	1	2
Substitution	G	A	A K G W I D S	A	A	A	A	A	A	G	A	A
Effect*	Ac	Ac	(An)	Ac	No	No	Ac	No	Ac	No	(An)	Ac
Wild-type	Y	R	Y	V	A	R	E	Q		S	C	R
							1	1		1	1	1
94-104	9	9	9	9	9	9	0	0		0	0	0
	4	5	6	7	8	9	0	1		2	3	4
Substitution	G	A	A	A	G	A	A	A		A	A	A
Effect*	Ac	(An)	(An)	Ac	Ac	Ac	(An)	No		Ac	Ac	Ac

Ac, Activation-inducing; *No,* nonstimulatory; (An) antagonistic.
*Effect of substituted peptide.

harvested and stored at −80° C before testing. Cytokine levels in the supernatants of activated T cells were determined by ELISA, with cytokine-specific antibodies for IL-2, IL-4, IFN-γ, and IL-5, as has been described.[31] The sensitivity of each assay was approximately 50 pg/ml.

Analysis of CD40L expression by flow cytometry

One million T cell cells per milliliter were stimulated with 2×10^6 irradiated (50 Gy), autologous EBV-LCL that had been pulsed for 3 hours with 1 μg/ml of specific immunogenic peptide in the presence or absence of a 100-fold molar excess of substituted peptide for 16 hours. To prevent downregulation of CD40L on activated T cells as a result of interaction with CD40 expressed on the APC, 1 μg/ml of anti-CD40L monoclonal antibody (mAb) was added to the cultures (R. Noelle, personal communication). Methods of immunofluorescence staining, flow cytometry, and data analysis have been described by Lanier and Recktenwald.[32] Flow cytometry to analyze the expression of CD40L on activated T-cell clones was performed with a FACS Calibur flow cytometer (Becton Dickinson Co., San Jose, Calif.) and the anti-CD40L mAb LL48 (CD154, kindly provided by Dr. J. Banchereau, Schering-Plough Research Laboratories, Dardilly, France).

Measurement of Ig production

One million T cells per milliliter were stimulated with 2×10^6 irradiated (50 Gy), autologous EBV-LCL (AR-142) or 2.5×10^5 HLA-matched monocytes (NP-44) that had been pulsed for 3 hours with 1 μg/ml of specific immunogenic peptide in the presence or absence of a 100-fold molar excess of substituted peptide, and culture supernatants were harvested after 24 hours of culture. The supernatants were added to 10^4 highly purified human B lymphocytes (> 99% $CD20^+$), isolated from normal spleens,[33] which were then stimulated with the anti-CD40 mAb 89[34] in 96-well round-bottomed plates (Flow) in Yssel's medium, supplemented with 10% fetal calf serum and 10 μg/ml ultrapure transferrin (Pierce, Rockford, N.Y.) in a final volume of 200 μl. For T cell–mediated IgE synthesis, T-cell clone NP-44 was activated for 24 hours with HLA-matched monocytes, pulsed for 3 hours with 1 μg/ml of p94-104 in the presence or absence of a 100-fold molar excess of Y96A, and after washing of the cells, 10^4 T cells were added to 10^4 purified B cells in the presence of exogenous rIL-4 (100 U/ml). Fourteen days later, the supernatants from triplicate cultures of B cells were harvested, pooled, and analyzed for Ig isotype content by ELISA as described previously.[33] The sensitivity of the ELISA was determined with a calibrated standard from Behring (Marburg, Germany) and was 0.2 ng/ml.

Generation of Der p 1–derived peptides

Der p 1–derived peptides, representing minimal T-cell activation–inducing epitopes or including single substitutions, were synthesized on an advanced Chemtech 350 Synthesizer (Advanced Chemtech, Louisville, Ky.). The carboxyl-terminal amino acid coupled to Wang resin was used as a the solid phase and the peptides were extended with FMOC chemistry, according to the manufacturer's instructions. Peptides were characterized by laser desorption mass spectrometry (FinniganMat, San Jose, Calif.), peptide sequencing, and quantitative amino acid analysis and had a purity of approximately 90%.

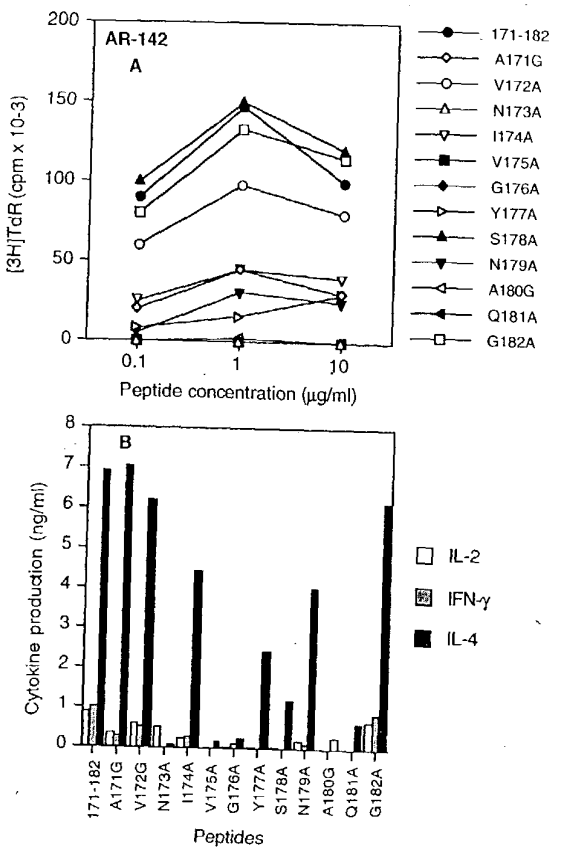


FIG. 1. Effect of amino acid substitutions in peptide p171-182 on proliferation and cytokine production of Der p 1–specific T-cell clone AR-142. T cells were stimulated with various concentrations of wild-type and substituted peptides, in presence of autologous EBV-LCL, and proliferative responses **(A)** were measured after 72 hours as described in Methods section. Production of IL-2, IL-4, and IFN-γ **(B)** was analyzed after incubation of T cells with 1 μg/ml of peptide and EBV-LCL for 48 hours.

RESULTS

Mutational analysis within minimal T-cell activation–inducing epitopes

The T-cell clones NP-12, NP-44, and AR-142 specifically proliferate and secrete cytokines in response to stimulation with the immunogenic, Der p 1–derived wild-type peptides p94-104 and p171-182, respectively, in the presence of autologous or HLA-matched APCs. To generate modified peptides that could block allergen-specific activation of these T-cell clones, single amino acid substitutions were introduced into the immunogenic peptide sequences (Table I). Stimulation of the T-cell clones with autologous EBV-LCL in the presence of various concentrations of substituted Der p 1–derived peptides resulted in various levels of T-cell proliferation as well as cytokine production. Maximal proliferative responses to wild-type peptide were obtained at approximately 1 μg/ml of peptide as shown in Fig. 1 and Fig. 2. Peptides derived from p171-182 with substitutions of amino acid residues at positions 173, 175, 176, 180, and 181 with Ala (A) or Gly (G) failed to induce proliferative responses and production of IL-2, IL-4, and IFN-γ by clone AR-142 (Fig. 1). In addition, peptides in which the neutral Asn (N) residue at position 173 was replaced by either basic [Lys:K], hydrophobic [Trp:W, Ile:I], acidic [Asp:D], or hydrophilic residue [Ser:S] also did not induce proliferation and cytokine production of AR-142 (results not shown), indicating that Asn at position 173 is one of the critical residues required for T-cell activation. Similarly, substitutions with Ala or Gly at positions 95, 96, 100, and 101 of peptide 94-104 resulted in modified peptides that were ineffective in inducing proliferation and IL-2, IL-4, and IFN-γ production by T-cell clones NP-12 and NP-44 (Fig. 2). The lack of proliferation-inducing capacity of these peptides was always associated with the inability to induce cytokine production, and no dissociation between proliferation and cytokine production or between the production of IL-2, IL-4, and IFN-γ was observed. Substitutions at amino acid positions other than those mentioned above induced normal or only slightly reduced proliferative responses and cytokine production of the T-cell clones compared with those induced with the native, nonsubstituted, peptides.

Inhibition of proliferation by antagonistic peptides

Next we determined whether the substituted peptides that failed to induce T-cell activation could act as antagonists by inhibiting the wild-type peptide–induced proliferation of the Der p 1–specific T-cell clones. Stimulation of T-cell clone AR-142 with wild-type peptide p171-182 and a 10-fold molar excess of the substituted peptides N173A or Q181A cell clone AR-142 resulted in an inhibition by more than 50% of the proliferation induced by p171-182 only (Fig. 3). Maximal inhibition of proliferative responses of approximately 70% was observed with a 100-fold molar excess of antagonistic peptides. Three nonstimulatory, modified peptides, generated by single amino acid substitutions in the p94-104 peptide, had strong inhibitory activities on the proliferation of T-cell clones NP-12 and NP-44. Substituted peptides R95A, Y96A, and Q101A strongly inhibited the proliferation of T-cell clone NP-12 in a dose-dependent manner. Weaker inhibitory effects of these peptides were observed on proliferative responses of T-cell clone NP-44. Preincubation of the APCs for 6 hours with a 100-fold molar excess of antagonistic peptides, before addition of the wild-type peptide, yielded similar inhibitory effects (S. Fasler, unpublished results).

The substituted peptides that inhibited Der p 1–induced proliferation of T-cell clone AR-142 were not able to block proliferative responses of T-cell clones NP-12 and NP-44 and vice versa, indicating that the inhibitory effect of these peptides was not due to toxic effects. All other substituted peptides with stimulatory activity, listed in Table I, did not affect Der p 1–specific proliferative responses of the T-cell clones at any concentration used (data not shown). Further mutational analysis

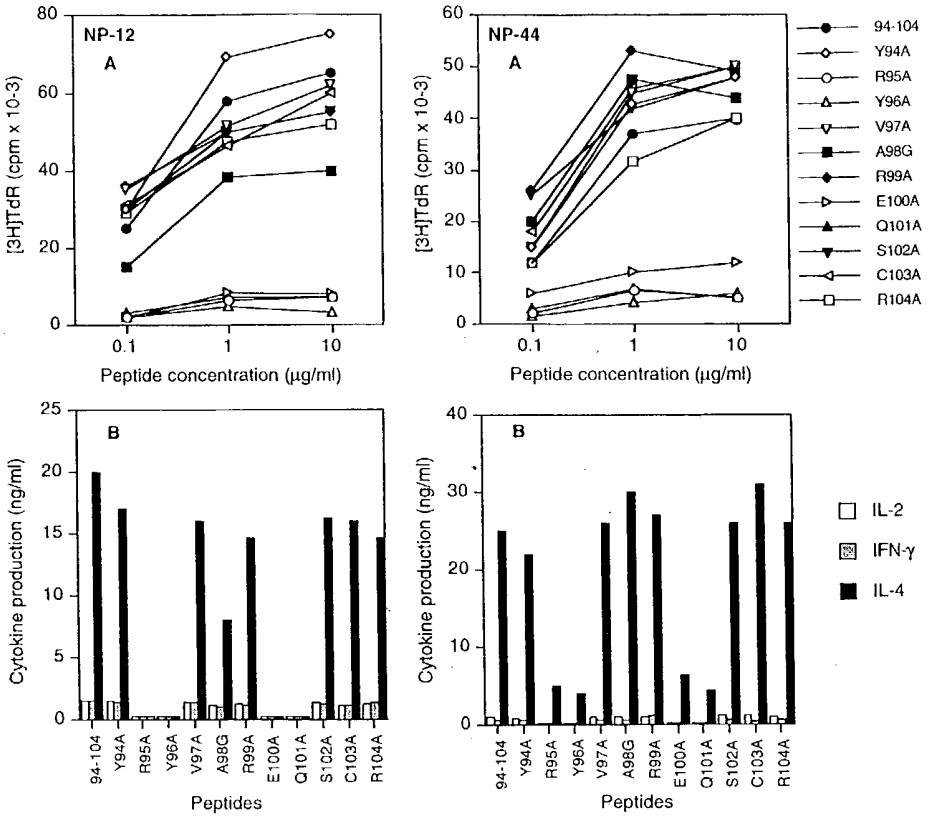


FIG. 2. Effect of amino acid substitutions in peptide p94-104 on proliferation and cytokine production of Der p 1–specific T-cell clones NP-12 and NP-44. Proliferative responses **(A)** and cytokine production **(B)** of T-cell clones NP-12 and NP-44 were measured as described in Fig. 1.

of residue 173 in the wild-type peptide p171-182, after substitution of Asn with Lys, Gly, Trp, Ile, Asp, and Ser, produced peptides with varying antagonistic activities. The inhibition of the Der p 1–induced proliferation of T-cell clone AR-142, at a 100-fold molar excess of substituted peptide, ranged from 9% to 78% with decreasing inhibitory activity, in the order Ala, Ser, Gly, Lys, Trp, Ile, and Asp (data not shown).

Inhibition of cytokine production by antagonistic peptides

To further examine the effects of substituted peptides on T-cell effector function, the cytokine production by Der p 1–specific T-cell clones, induced by the wild-type peptides in the presence of an excess amount of modified peptide, was studied. As is shown in Table II, the antagonistic peptides inhibited the proliferative responses of the Der p 1–specific T-cell clones in a dose-dependent fashion with a maximal inhibitory effect at a 100-fold molar excess. Therefore this concentration of peptide was used in further experiments. The antagonistic peptides R95A, Y96A, and E100A, which inhibited the proliferation of T-cell clones NP-12 and NP-44, also strongly blocked the production of IL-2 and IFN-γ at a 100-fold molar excess. Comparable results were obtained with the antagonistic peptides N173S, Q181A, and N173A, which had strong inhibitory activities on the proliferative responses of T-cell clone AR-142. However, the relatively high production levels of IL-4 and IL-5, in particular those produced by T-cell clones NP-44 and AR-142, were only partially affected (Table II), even in the presence of a 500-fold molar excess of antagonistic peptide (data not shown), suggesting a difference in sensitivity of the T_{H1} versus the T_{H2} cytokines to the inhibitory effects of the antagonistic peptide. However, none of the substituted peptides had a differential effect on the production of the various cytokines; furthermore, antagonistic peptides that inhibited cytokine production also inhibited proliferative responses.

Antagonistic peptides inhibit help for IgE synthesis by B cells by blocking IL-4 production and expression of CD40L

Activated Der p 1–specific T-cell clones give help for IgE synthesis to purified B cells for the synthesis of IgE through their production of IL-4, as well as a signal mediated by the interaction of CD40 and its ligand CD40L. In view of their inhibitory effects on the cytokine production of Der p 1–specific T-cell clones, we investi-

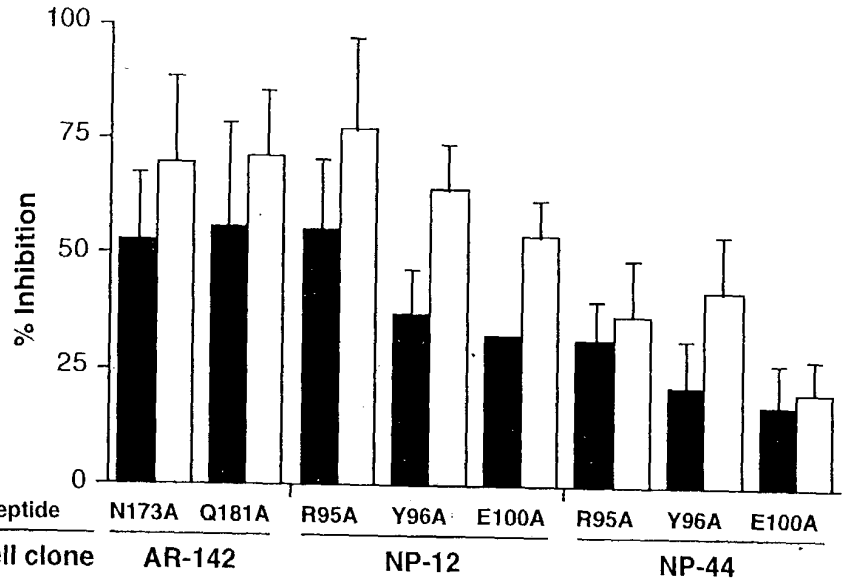


FIG. 3. Inhibition of proliferative responses of Der p 1–specific T-cell clones by antagonistic *Der* p 1–derived peptides. T-cell clones AR-142, NP-12, and NP-44 were incubated with autologous EBV-LCL in the presence of 0.1 μg/ml of the native peptides p171-182 or p94-104, respectively, and a 10-fold *(filled columns)* or 100-fold *(open columns)* molar excess of antagonistic peptides. Proliferative responses were measured after 72 hours as described in Methods section. Results are expressed as percentage inhibition of native peptide–induced proliferation and represent mean and standard deviation of three independent experiments.

TABLE II. Inhibition of cytokine production of Der p 1–specific T-cell clones by substituted Der p 1–derived peptides

T-cell clone*	Substituted peptide	% Inhibition of cytokine production†			
		IFN-γ	IL-2	IL-4	IL-5
AR-142	Q181A	86	100	51	34
	N173A	93	100	50	37
	N173S	94	100	50	29
	N173G	76	100	44	11
	N173K	68	100	29	4
	N173I	63	89	27	5
	N173W	56	73	28	3
	N173D	37	12	6	−12
NP-12	R95A	44	99	69	81
	Y96A	70	95	75	92
	E100A	90	100	86	99
NP-44	R95A	98	100	46	50
	Y96A	101	100	44	60
	E100A	98	100	43	78

*T cells were stimulated with autologous EBV-LCL, pulsed with 1 μg/ml of wild-type peptides p171-182 (AR-142) or p94-104 (NP-12 and NP-44), respectively, and a 100-fold molar excess of substituted peptides as described in Methods section.

†Cytokine production was measured after 48 hours of culture as described in Methods section. Values for cytokine production after stimulation with wild-type peptides are similar to those shown in Figs. 1 and 2.

gated whether antagonistic peptides could inhibit T-cell–mediated induction of IgE synthesis by purified B cells in vitro. As demonstrated previously, the addition of culture supernatants of T-cell clone NP-44, activated with peptide p94-104 and APC, to highly purified B cells resulted in the induction of high levels of polyclonal IgE (Fig. 4).[34] Supernatants collected from cultures of NP-44 cells, stimulated with the wild-type peptide p94-104 and a 100-fold molar excess of Y96A, showed a reduction of approximately 50% to induce IgE synthesis in vitro by B cells preactivated with anti-CD40 mAb (Fig. 4, *A*). Similar results were obtained with T-cell clone AR-142 and N173A (Fig. 4, *B*), reflecting the level of partial inhibition on the production of IL-4 by these antagonistic peptides. To determine whether such antagonistic peptides could also block IgE synthesis mediated by activated T-cell clones in vitro, purified B cells were cocultured with NP-44 cells, stimulated with p94-104, a 100-fold molar excess of substituted peptide Y96A or combinations of p94-104 and Y96A, and IgE production was analyzed after 12 days of culture. As shown in Fig. 4, *C*, NP-44 cells, stimulated with p94-104 and excess Y96A, were no longer able to provide help to B cells for the production of IgE, even in the presence of exogenous IL-4. To examine whether the presence of antagonistic peptides can interfere with T cell–B cell–mediated signals leading to IgE synthesis, their effect on the induction of CD40L expression was analyzed. T-cell clones NP-44 and AR-142 expressed significant levels of CD40L, as detected by flow cytometry, 6 hours after stimulation of the cells with APCs and wild-type peptide. As shown in Fig. 5, the presence of a 100-fold molar excess of Y96A or N173A during antigen-specific stimulation of T-cell clones NP-44 and AR-142 strongly inhibited the induction of CD40L expression. Taken together, these results suggest that antagonistic peptides

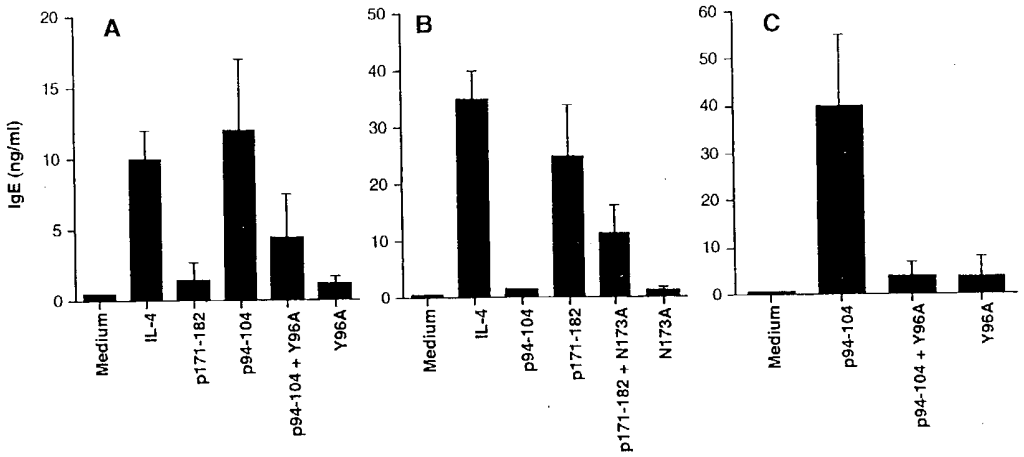


FIG. 4. Inhibition of T-cell help for IgE synthesis by antagonistic peptides. Ten thousand highly purified B cells (> 99%) were incubated with anti-CD40 antibody (1 μg/ml) and cultured with 25% dilution of culture supernatants from T-cell clones NP-44 **(A)** and AR-142 **(B)** that had been stimulated with wild-type peptide (1 μg/ml), APCs, and a 100-fold excess of antagonistic peptides for 24 hours as described in Methods section. Alternatively, 10^4 NP-44 cells that had been stimulated as described in **A** were washed twice with medium cultured with highly purified B cells in presence of 100 U/ml rIL-4 **(C)**. After 14 days, IgE levels in B-cell cultures were measured by ELISA as described in Methods section.

block Ig isotype switching and subsequent IgE production by B cells in vitro by decreasing IL-4 production by Der p 1–specific T-cell clones and by interfering with the expression of the CD40L and possibly other surface molecules on T cells, required for optimal T cell–B cell interaction.

DISCUSSION

In this study we demonstrate that certain modified peptide sequences, derived from native T-cell activation–inducing sequences in the major house dust mite allergen Der p 1, fail to induce proliferative responses and cytokine production of Der p 1–specific T_{H2}-like clones. In addition, when added in excess, they strongly inhibit allergen-induced activation, resulting in the inability of these T-cell clones to give help to B cells for the synthesis of IgE. An overview of the effects of the substituted peptides on T-cell activation is shown in Table I. These peptides were derived from the p94-104 and p171-182 sequences, which are minimal T-cell activation–inducing epitopes on Der p 1 and are recognized in the context of HLA DR11 and HLA-DR15, respectively. Single substitutions of Arg^{95}, Tyr^{96}, Glu^{100}, and Gln^{101} in p94-104 abolished proliferative responses of T-cell clones NP-12 and NP-44. Similarly, stimulation of T-cell clone AR-142 with modified peptide p171-182 containing substitutions at Asn^{173}, Val^{175}, Gly^{176}, Ser^{178}, Ala^{180}, and Gly^{181} by Ala did not induce a proliferative response when used at similar concentrations as the nonsubstituted peptide. In addition, stimulation of the T-cell clones with these modified peptides also failed to induce the production of cytokines. These results indicate that the above-mentioned residues are critical for recognition of p94-104 and p171-182 by the TCR or may function as anchors involved in binding to agretopes on HLA DRB1*1101 and DRB1*1501, respectively. The failure of some of the substituted peptides to induce T-cell activation might be due to an impaired interaction of these peptides with the TCR, with HLA-DR molecules, or both. Future studies with MHC-binding analysis may provide an answer to this question.

The Der p 1–specific T-cell clones used in this study have a T_{H2}-like cytokine production profile, secreting high levels of IL-4 and IL-5 and low levels of IFN-γ upon activation. The production of the latter cytokine probably is due to the presence of IL-12 in the cultures, which has been shown to induce the production IFN-γ by polarized T_{H2} cells.[35, 36] Furthermore, IL-2, which is produced by human T_{H1} as well as T_{H2} clones, was produced at detectable levels. Stimulation of the T-cell clones with the antagonistic peptides resulted in a general inhibition of activation and did not induce a segregation of proliferation and cytokine production, as has been reported for murine[21, 23] and certain human T_{H0} clones.[25] In addition, incubation of the Der p 1–specific T-cell clones with the modified peptides, at concentrations required for optimal T-cell proliferation, did not induce a state of unresponsiveness as shown previously for altered peptides derived from ovalbumin.[37] Our results also partially contrast with those of Tsitoura et al.[26] and Ikagawa et al.,[27] who reported that stimulation of allergen-specific T-cell clones with altered peptides resulted in enhanced production of IFN-γ but did not affect the production of IL-4. In view of the latter studies showing that single amino substitutions can alter the profile of cytokine production by T-cell clones, the effects of amino acid residues other than the Ala substitutions were studied in the antagonistic peptide N173A. The initial substitution with the neutral amino acid residue, Ala or Ser, produced the most efficient antago-

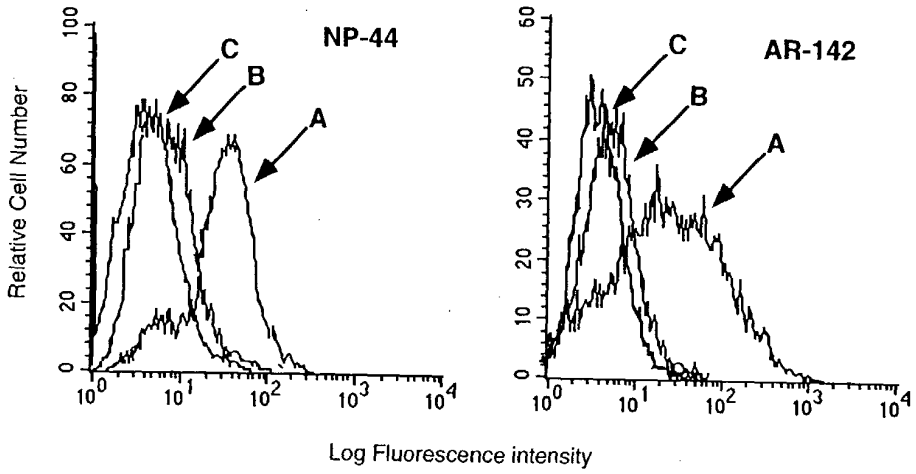


FIG. 5. Inhibition of CD40L expression by antagonistic peptides. NP-44 and AR-142 cells with autologous EBV-LCL and wild-type peptides p94-104 or p171-182, respectively, in absence *(A)* or presence *(B)* of a 100-fold molar excess of antagonistic peptide as described in Methods section. CD40L expression was analyzed by flow cytometry after 6 hours of culture. The *x*- and *y*-axes represent fluorescence (four-decade log scale) and relative cell number, respectively. CD40L expression on resting, nonstimulated T-cell clones is shown in histogram *(C)*.

nists, whereas substitution with the hydrophobic Ile or acidic Asp was less effective in inhibiting cytokine production (data not shown). However, none of these other substitutions at position 173, albeit resulting in the generation of peptides with antagonistic activity, selectively inhibited cytokine production or proliferation alone or changed the T_{H2}-like cytokine secretion pattern of the T-cell clones.

Among the substituted peptides that failed to induce activation of the Der p 1–specific T-cell clones, three of four derived from p94-104 and two of six derived from p171-182 were found to have strong antagonistic activity (Tables I and II and Fig. 3). Incubation of the T-cell clones in the presence of the nonsubstituted immunogenic peptide in the presence of an excess of these antagonistic peptides resulted in a strong inhibition of proliferation as well as of the production of IL-2 and IFN-γ. Because the proliferative responses of the Der p 1–specific T-cell clones used in this study are strictly dependent on the presence of IL-2 despite their T_{H2}-like phenotype (H. Yssel, unpublished results), the inhibitory effects of the antagonistic peptides are therefore likely to be the result of the strong inhibition of IL-2 production as well as of CD25 expression (data not shown). Interestingly, however, the TCR antagonists had less effect on IL-4 and IL-5 production induced by the nonsubstituted native peptides. For example, peptide analogues N173A and Q181, derived from the p171-182 molecule, strongly inhibited proliferative responses and IL-2 production already at a 10-fold molar excess over the native peptide, whereas the presence of a 100-fold molar excess of these antagonists resulted in the decrease of IL-4 and IL-5 production of less than 50%. Although these results do not show an uncoupling between proliferation and cytokine production, they do not rule out the possibility that the antagonistic peptides may have differential effects on signal transduction by the TCR. In particular, the low inhibitory effects of the antagonistic peptides on IL-4 production are in line with previously published observations, suggesting that the synthesis of IL-4 is less dependent on costimulatory signals, as demonstrated by the ability of mouse T_{H2} clones to continue to produce IL-4 when stimulated with chemically modified APCs.[21] Moreover, IL-4 production seems to be favored by signals resulting from less potent TCR-ligand interactions, such as those after stimulation of T cells with peptides that have a low affinity for the TCR or for MHC class II molecules.[21] In addition, it has been reported that mouse T_{H1} and T_{H2} clones use different T-cell signaling pathways, suggesting that TCR-mediated signaling is likely to differ between IL-4 and IFN-γ production.[37] On the other hand, however, because the antagonistic peptides, when used by themselves had an "all or none" effect on the cytokine production profile of the Der p 1–specific T-cell clones, it cannot be excluded that the observed differences in sensitivity between the various cytokines, after stimulation of the cells in the presence of antagonist and native peptide, may be the result of differences in the threshold of the activation signal or signals mediated by the TCR, rather than differences in signaling pathways.

The synthesis of IgE by B cells has been shown to be specifically induced by IL-4 and IL-13 in the presence of a signal provided for by the interaction of CD40, expressed on B cells and CD40L, expressed on activated T cells.[4-9] Moreover, IL-4/IL-13-induced IgE synthesis is inhibited by IFN-γ.[4] Culture supernatants of T-cell clones NP-44 and AR-142 that had been activated with specific peptide in the presence of excess antagonistic peptides Y96A and N173A, respectively, were still able to induce anti–CD40-driven

IgE synthesis by B cells. However, levels of IgE synthesis were inhibited by approximately 50% to 75%, compared with those induced by T-cell supernatants generated after stimulation of the T-cell clones with native peptides alone. These results reflect the partial inhibitory effects on the reduced levels of IL-4 and, most likely, IL-13 production by the T-cell clones. In addition, T-cell clone NP-44, activated in the presence of native peptide p94-104 and excess antagonistic peptide Y96A, was no longer able to give help to B cells for the synthesis of IgE, even in the presence of exogenous IL-4. The inhibitory effects on these antagonistic peptides on T cell–mediated IgE synthesis appear to be the result of their capacity to downregulate or prevent the induction of CD40L expression on the T cells. Because these results underscore the notion that the antagonistic peptides used in this study do not appear to induce uncoupling of TCR-mediated signals leading to proliferation, cytokine synthesis, and expression of cell surface molecules, it is likely that the expression of other cell surface molecules that may be associated with productive T cell–B cell interaction, such as membrane-bound tumor necrosis factor (TNF)-α,[33] are downregulated as well. Taken together, these results demonstrate that antagonistic Der p 1–derived peptides are able to effectively inhibit T_{H2} cell-mediated IgE synthesis by decreasing the production of the IgE synthesis-inducing cytokine IL-4, as well as by inhibiting the functional expression of cell surface molecules, which are important for productive T cell–B cell interactions.

In conclusion, in this study we describe the generation of antagonistic peptides, after the substitution of single amino acid residues that are effectively interfering with the activation of and help for IgE synthesis by Der p 1–specific T-cell clones. Although the activation of the T-cell clones in this study was carried out with wild-type Der p 1–derived peptides, identical inhibitory effects of the antagonistic peptides were also observed when the T-cell clones had been activated with the native Der p 1 allergen (data not shown). In this respect, these peptides may be of potential interest for therapeutic use. It must be noted, however, that house dust mite extracts consist of a large number of major allergens, each containing multiple T-cell activation–inducing epitopes.[15, 38, 39] In addition, allergens are recognized in the context of different HLA class II molecules, which hampers efforts to design immunotherapy protocols with the aim of inactivating allergen-specific T cells. It is therefore important to state that despite their effectiveness in inhibiting IgE synthesis in vitro, more information is required to evaluate the potential use of antagonistic allergen-derived peptides in immunotherapy.

The authors would like to thank Patricia Larenas, Jean Chang, and Bruce Bennet for expert technical assistance and JoAnn Katheiser for secretarial help. The DNAX Research Institute for Molecular and Cellular Biology is supported by Schering-Plough Corporation.

REFERENCES

1. Chapman MD, Heymann PW, Platts-Mills TA. Epitope mapping of two major inhalant allergens, *Der p* I and *Der f* I, from mites of the genus Dermatophagoides. J Immunol 1987;139:1479-4.
2. Chua KY, Kehal PK, Thomas WR. Sequence polymorphisms of cDNA clones encoding the mite allergen *Der p* I. Int Arch Allergy Immunol 1993;101:364-8.
3. Heymann PW, Chapman MD, Aalberse RC, Fox JW, Platts-Mills TA. Antigenic and structural analysis of group II allergens (*Der f* II and *Der p* II) from house dust mites (Dermatophagoides spp). J Allergy Clin Immunol 1989;83:1055-7.
4. Pène J, Rousset F, Brière F, Chrétien I, Bonnefoy JY, Spits H, et al. IgE production by normal human lymphocytes is induced by interleukin (IL)-4 and suppressed by interferon (IFN)-γ and -α, and prostaglandin E2. Proc Natl Acad Sci U S A 1988;85:6880-4.
5. Punnonen J, Aversa G, Cocks BG, Mckenzie ANJ, Menon S, Zurawski G, et al. Interleukin 13 induces interleukin 4-independent IgG4 and IgE synthesis and CD23 expression by human B cells. Proc Natl Acad Sci U S A 1993;90:3730-4.
6. Punnonen J, de Vries JE. IL-13 induces proliferation, Ig isotype switching, and Ig synthesis by immature human fetal B cells. J Immunol 1994;152:1094-2.
7. Jabara HH, Fu SM, Geha RS, Vercelli D. CD40 and IgE: synergism between anti-CD40 monoclonal antibody and interleukin 4 in the induction of IgE synthesis by highly purified human B cells. J Exp Med 1990;172:1861-4.
8. Hollenbaugh D, Grosmaire LS, Kullas CD, Chalupny NJ, Braesch-Andersen S, Noelle RJ, et al. The human T cell antigen gp39, a member of the TNF gene family, is a ligand for the CD40 receptor: expression of a soluble form of gp39 with B cell costilmulatory activity. EMBO J 1992;11:4313-1.
9. Spriggs MK, Armitage RJ, Strockbine L, et al. Recombinant human CD40 ligand stimulates B cell proliferation and Immunoglobulin E secretion. J Exp Med 1992;176:1543-50.
10. Allen RC, Armitage RJ, Conley ME, Rosenblatt H, Jenkins NA, Copeland NG, et al. CD40 ligand gene defects responsible for X-linked hyper-IgM syndrome. Science 1993;259:990-4.
11. Aruffo A, Farrington M, Hollenbaugh D, Li X, Milatovich A, Nonoyama S, et al. The CD40 ligand, gp39, is defective in activated T cells from patients with X-linked hyper-IgM syndrome. Cell 1993;72:291-7.
12. Di Santo JP, Bonnefoy JY, Gauchat JF, Fischer A, De Saint Basile G. CD40 ligand mutations in X-linked immunodeficiency with hyper-IgM. Nature 1993;361:541-3.
13. Fuleihan R, Ramesh N, Loh R, Jabara H, Rosen RS, Chatila T, et al. Defective expression of the CD40 ligand in X chromosome-linked immunoglobulin deficiency with normal or elevated IgM. Proc Natl Acad Sci U S A 1993;90:2170-4.
14. Wierenga EA, Snoek M, de Groot C, Chrétien I, Bos JD, Jansen HM, et al. Evidence for compartmentalization of functional subsets of $CD4^+$ T lymphocytes in atopic patients. J Immunol 1990;144: 4651-6.
15. Yssel H, Johnson KE, Schneider PV, Wideman J, Terr A, Kastelein R, et al. T cell activation-inducing epitopes of the house dust mite allergen *Der p* I: induction of a restricted cytokine production profile of *Der p* I-specific T cell clones upon antigen-specific activation. J Immunol 1992;148:738-45.
16. de Waal Malefyt R, Abrams JS, Zurawski SM, Lecron J-C, Mohan-Peterson S, Sanjanwala B, et al. Differential regulation of IL-4 and IL-13 expression by human T cell clones and EBV-transformed B cells. Int Immunol 1995;7:1405-16.
17. Kim J, Urban RG, Strominger JL, Wiley DC. Toxic shock syndrome toxin-1 complexed with a class II major histocompatibility molecule HLA-DR1. Science 1994;266:1870-4.
18. Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 1993;364: 33-9.
19. Alexander J, Snoke K, Ruppert J, Sidney J, Wall M, Southwood S, et al. Functional consequences of engagement of the T cell receptor by low affinity ligands. J Immunol 1993;150:1-7.

20. Alexander J, Ruppert J, Snoke K, Sette A. TCR antagonism and T cell tolerance can be independently induced in a DR-restricted, hemagglutinin-specific T cell clone. Int Immunol 1994;6:363-7.
21. Evavold BD, Allen PM. Separation of IL-4 production from Th cell proliferation by an altered T cell receptor ligand. Science 1991;252: 1308-10.
22. Evavold BD, Williams SG, Hsu BL, Buus S, Allen PM. Complete dissection of the Hb(64-76) determinant using T helper 1, T helper 2 clones, and T cell hybridomas. J Immunol 1992;148:347-53.
23. Evavold BD, Sloan-Lancaster J, Hsu BL, Allen PM. Separation of T helper 1 clone cytolysis from proliferation and lymphokine production using analog peptides. J Immunol 1993;150:3131-40.
24. De Magistris MT, Alexander J, Coggeshall M, Altman A, Gaeta FC, Grey HM, et al. Antigen analog-major histocompatibility complexes act as antagonists of the T cell receptor. Cell 1992;68:625-34.
25. Windhagen A, Scholz C, Hollsberg P, Fukaura H, Sette A, Hafler DA. Modulation of Cytokine patterns of human autoreactive T cell clones by a single amino acid substitution of their peptide ligand. Immunity 1995;2:373-80.
26. Tsitoura DC, Holter W, Cerwenk A, Gelder CM, Lamb JR. Induction of anergy in human T helper 0 cells by stimulation with altered T cell antigen receptor ligands. J Immunol 1996;156:2801-8.
27. Ikagawa S, Matsushita S, Chen YZ, Ishikawa T, Nishimura Y. Single amino acid substitutions on a Japanese cedar pollen allergen (*Cry j* 1)-derived peptide induced alterations in human T cell responses and T cell receptor antagonism. J Allergy Clin Immunol 1996;97:53-64.
28. Spits H, Yssel H, Terhorst C, de Vries JE. Establishment of human T lymphocyte clones highly cytotoxic for an EBV-transformed B cell line in serum-free medium: isolation of clones that differ in phenotype and specificity. J Immunol 1982;128:95-105.
29. Yssel H, de Vries JE, Koken M, Van Blitterswijk W, Spits H. Serum free medium for the generation and propagation of functional cytotoxic and helper T cell clones. J Immunol Methods 1984;72:219-27.
30. Figdor CG, Bont WS, Touw I, Roos D, Roosnek EE, de Vries JE. Isolation of functionally different human monocytes by counterflow centrifugation elutriation. Blood 1982;60:46-53.
31. Abrams JS. Immunoenzymatric assay of mouse and human cytokines using NIP-labeled anti-cytokine antibodies. Curr Protocol Immunol 1994;6.20.1A.
32. Lanier LL, Recktenwald DJ. Multicolor immunofluorescence and flow cytometry. Methods 1991;2:192-9.
33. Aversa G, Punnonen J, de Vries JE. The 26-kD transmembrane form of tumor necrosis factor alpha on activated CD4+ T cell clones provides a costimulatory signal for human B cell activation. J Exp Med 1993;177:1757-85.
34. Fasler S, Aversa G, Terr A, Thestrup-Pedersen K, de Vries JE, Yssel H. Induction of anergy in allergen-specific Th2 cells by allergen-derived peptides results in lack of B-cell help for IgE synthesis: reversal by IL-2, but not by IL-4 or IL-13. J Immunol 1995;155:4199-206.
35. Manetti R, Gerosa F, Giudizi MG, Biagiotti R, Parronchi P, Piccinni M-P, et al. Interleukin 12 induces stable priming for interferon γ (IFN-γ) production during differentiation of human T helper (Th) cells and transient IFN-γ production in established Th2 cell clones. J Exp Med 1994;179:1273-83.
36. Yssel H, Fasler S, de Vries JE, de Waal Malefyt R. Interleukin 12 (IL-12) transiently induces IFN-γ transcription and protein synthesis in human CD4+ allergen-specific helper type 2 (Th2) T cell clones. Int Immunol 1994;6:1091-6.
37. Sloan-Lancaster J, Evavold BD, Allen PM. Induction of T-cell anergy by altered T-cell-receptor ligand on live antigen-presenting cells. Nature 1993;363:156-9.
38. van Neerven JR, van t'Hof W, Ringrose JH, Jansen HM, Aalberse RC, Wierenga EA, et al. T cell epitopes of house dust mite major allergen *Der p* II. J Immunol 1993;151:2326-35.
39. Higgins JA, Thorpe CJ, Hayball JD, O'Hehir RE, Lamb JR. Overlapping T-cell epitopes in the group I allergen of Dermatophagoides species restricted by HLA-DP and HLA-DR class II molecules. J Allergy Clin Immunol 1994;93:891-9.

5/1870

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

NPL ___ Adonis ___
MIC ___ BioTech ___ MAIN ___
NO X Vol NO ___ NOS ___
Ck Cite ___ Dupl Request ___
Call # 9122 WK

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al,* Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al,* Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al,* A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

15093866

Therapeutic Drug Monitoring
15:514–520 © 1993 Raven Press, Ltd., New York

Biotechnology: An Introduction to Recombinant DNA Technology and Product Availability

†‡Ronald P. Evens and *Mark Witcher

**Manufacturing Department and †Professional Services Department, Amgen Inc., Thousand Oaks; and ‡School of Pharmacy, University of Southern California, Los Angeles, California, U.S.A.*

Summary: Exciting advances in biotechnology have led to the development of innovative biological products to improve health care. In this review article, the major techniques of product development and manufacturing in biotechnology are discussed, with a focus on recombinant DNA technology. In addition, monoclonal antibody, nucleotide blockade, polymerase chain reaction, antisense, and gene therapy technologies will be defined briefly. For recombinant-DNA technology, the issues of gene isolation, gene cloning, protein expression, scale-up (manufacturing), and quality assurance are addressed. The 16 approved products and the research pipeline are characterized. In addition, major usage issues for biological products are noted. This review serves as an introduction to the science and applications of biotechnology present and future. **Key Words:** Biotechnology—Recombinant DNA technology—Biological products.

During the past two decades, biotechnology has evolved from a laboratory science to a dynamic research-based business and an established form of therapy that benefits patients everyday. The sophisticated science of biotechnology continues to unfold and is expanding rapidly, from the core technologies of recombinant DNA procedures and monoclonal antibodies to whole new techniques, such as antisense (anti-RNA code blockade) (1). The biotechnology revolution is a success because of the scientific ingenuity and entrepreneurial spirit of molecular biologists, biochemists, and investment capitalists, who have formed >1,000 biotechnology research companies in the United States that develop biological products to mitigate the effects of diseases in new ways.

Two indispensable factors in this success have been substantial research collaboration with universities and research and regulatory cooperation with government agencies. A 1991 survey showed that 742 biotechnology firms had ~$12 billion in revenue (2). In 1992, the top 79 biotechnology companies, according to stock market analysts, generated $2.127 billion in revenue (3), which was corroborated in a recent review article (13 molecules were responsible for $2.14 billion in sales in 1992) (4). The research and development commitment is impressive as well; 742 companies (1991 survey) reportedly invested >$3.63 billion in research (~30% of revenues) (2). From 1990 to the year 2000, the U.S. biotechnology market is projected to grow tenfold, including pharmaceutical, diagnostic, agricultural, and environmental products. This review article focuses on the scientific principles of biotechnology, product development issues, and products currently in use and under development.

DEFINITIONS

The molecules or biological products created by biotechnology research and manufacturing to date

Address correspondence and reprint requests to Dr. Ronald P. Evens at Amgen Inc., 1840 De Havilland Drive, Thousand Oaks, CA 91320, U.S.A.

are all proteins. The manufacturing process for these molecules uses a living system, usually a cell; this cell produces all of its normal proteins plus one additional protein, very similar to naturally occurring human molecule. Van Noordwijk offered a broadly applicable, yet simple definition of biotechnology: "the use of tissue cultures, living cells or cell enzymes to make a defined product" (5). More precise definitions exist and often delineate specific processes that make up biotechnology, such as hybridoma, cell fusion, and recombinant DNA. Furthermore, biotechnology can be described as a composite of several key scientific disciplines—molecular biology, microbiology, biochemistry, immunology, genetics, and engineering.

In the brief time period from 1982 to 1993, biological products have become widely available (15 molecules and 21 products by the fall of 1993). These products are being used in hundreds of thousands of patients annually for such common maladies as diabetes, cancer, heart attacks, and anemia. Biotherapy is a term now used to characterize the clinical use of these products. Its definition encompasses the administration of proteins, which often are duplicates or functional equivalents of naturally occurring molecules in the human body, in order to enhance normal body functions or replace deficient or dysfunctional proteins. Biotherapy is being integrated into patient care along with the established therapies of drugs, diet, surgery, physical procedures, radiation, and psychotherapy.

DESCRIPTION OF RECOMBINANT DNA TECHNOLOGY

Molecular biology is based on the central tenet that DNA makes RNA makes protein (6). This process comprises three major steps: transcription, translation, and completion of the protein. Genes are DNA nucleotide sequences, the basic units of heredity, and the starting material for recombinant DNA (r-DNA) technology. In the 23 pairs of chromosomes in the nucleus, made up of three to four billion base pairs of nucleotides, we need to locate the one gene of 50,000 genes responsible for producing a specific protein. In addition, genes contain coding DNA sequences (exons) responsible for protein production; a leader DNA sequence to initiate transcription; noncoding DNA material (introns), which must be eliminated in the formation of messenger RNA (m-RNA); and nucleotide sequences that indicate the terminal end of a gene and turn off transcription. The end product of r-DNA technology is a protein, which includes a set of amino acids of the correct number, type, and sequence. Several structural requirements are also critical for optimal activity of proteins; these include disulfide cross-linkage, three-dimensional folding, and glycosylation with sugar moieties of the right number, location, and type.

The steps in the process and the basic elements of recombinant DNA technology are outlined in Table 1 and are described herein. Recombinant DNA technology starts with the step of gene isolation, which includes three methods. One method begins with protein analysis; the composition of the protein must be identified fully with regard to the type and sequence of amino acids. Because the DNA codons (triplets) for each amino acid already are known, the DNA sequence then is synthesized in reverse from the protein. A second method involves isolation of the m-RNA that is responsible for producing a specific protein. Then, the viral enzyme, reverse transcriptase, is used to create the complementary DNA for the m-RNA. A third method is a screening process of the full genomic library when the protein cannot be identified fully. From two

TABLE 1. *Steps in process and basic elements of recombinant DNA technology*

Step 1
Gene isolation (three alternatives with different starting material)
From protein analysis to DNA triplets to DNA synthesis
From m-RNA identification plus viral reverse transcriptase to c-DNA formation
From chromosome pool with DNA probes to DNA screening for gene
Step 2
Cloning/expression (from genes to proteins in the laboratory)
Plasmid and gene-starting DNA materials
Restrictive endonuclease enzyme for plasmid opening
DNA ligase enzyme for plasmid closure with human gene
Recombinant plasmid entry into host cell
Host cell/plasmid cloning (duplication)
Gene expression of protein by host cells
Step 3
Scale-up process (manufacturing of commercial product)
Inoculum—daughter cells from master cell bank
Cell culture—serial fermentation—crude protein mixture
Purification—desired protein in bulk
Formulation—final product
Step 4
Quality control measures (tests of products and process)
Plasmid/host cell—integrity assessment
Bulk product—structure, potency, purity, activity
Process validation—efficiency and reproducibility
Final product—for bulk product, plus stability, sterility

separate peptide segments of the protein, two distinct oligonucleotide probes matching the two peptides are created. Next, the chromosome is denatured into several hundreds of thousands of DNA segments. This pool of DNA segments is searched using each oligonucleotide probe. A DNA segment that matches both probes would be a likely clone representing the target gene.

The second step in product development for r-DNA technology is cloning and expression, which is worked out in the laboratory. Cloning is duplication of a gene and its host cells. Expression is production of the protein from the gene in a host cell. The components of this step include the gene, plasmids, restriction endonucleases, DNA ligases, DNA linkers, DNA promoters/enhancers, and host cells.

Plasmids are circular rings of DNA that serve as the vector to carry the human gene into host cells. Key properties of plasmids are that they are DNA material, they have the capacity for autonomous replication of DNA, they allow ready acceptance of genes, they are transferable between cells, and they are stable during the recombinant process. Their source is bacteria, where they are responsible for transference of virulence or resistance, e.g., the pBR322 plasmid. Restriction endonuclease enzymes are employed to cut open plasmids at specific nucleotide triplets; a certain enzyme is chosen to open the plasmid at sites compatible with the terminal triplets of a gene. Bacteria again provide this tool, with ~100–200 distinct enzymes available.

DNA ligase enzyme is used to recombine the DNA material, incorporating the human gene into the bacterial plasmid. If the gene does not permit full closure of the plasmid ring, a DNA linker (segment of DNA) is added to piece together the gene with the plasmid ring. Sometimes a special DNA segment from a viral or bacterial source is added to the gene/plasmid DNA ring to increase gene replication and protein production. This recombined plasmid is then transferred to a host cell.

Host cells need to possess three primary characteristics: a short reproductive cycle, ease of growth in vitro, and acceptability of plasmids. Host cells can be fungal (e.g., yeast), bacterial (e.g., *E. coli*), or mammalian (e.g., Chinese hamster ovary) cells. The host cells with the recombined plasmids are placed in special nutrient mixtures designed for optimal cell viability and protein production. These cells are cloned (duplicated) along with intracellular plasmid multiplication. These host cells retain the capacity to produce all the various proteins in their normal life cycle and express one more protein, which is the duplicate or functional equivalent of the human protein produced by human gene expression. The cloning process results in a master working cell bank of special new parent cells capable of producing human proteins.

The site of manufacture of proteins in r-DNA technology is living cells. Scale-up of protein production is required to manufacture sufficient amounts of protein from very large cell colonies to meet patients' needs in a reasonably cost-effective manner. The manufacturing processes primarily are twofold, based on the type of host cells, that is, mammalian (eucaryote) or microbial (prokaryote) cells. In mammalian cell culture, cells grow slowly, with a doubling time of ~24–36 h. A production lot can require more than a month to produce. Prominent challenges are to maintain an environment that remains sterile, keeps cells viable and productive, and avoids any contamination over a 1–2 month period. Mammalian host cells secrete the proteins extracellularly, which enhances harvesting of proteins from media. Chinese hamster ovary cells are the mammalian cell line employed for epoetin alfa (EPOGEN) and alteplase (Activase). For microbial cell cultures, the cell doubling time can be as little as 30 min, which is highly desirable and, hence, is preferred for producing proteins. Microbial fermentation is an easier process and is more frequently used in r-DNA manufacturing. It is employed for filgrastim (NEUPOGEN). Microbial cells do not secrete proteins extracellularly and store proteins in vesicles, which complicates harvesting of proteins.

Four primary steps are involved in scale-up for mammalian r-DNA manufacturing: inoculum, cell culture, purification, and formulation. The inoculum step entails obtaining a single vial of daughter cells from the parent cells in the master working cell bank. These daughter cells are grown in vitro in small spinner flasks to initiate the manufacturing scale-up. This initial cell culturing expands cell volume with serially larger spinner flasks; cell density in the flasks is the critical factor in maximizing cell growth. The goal of this inoculum phase is to prepare a sufficient number of cells for conduct of the cell culture phase. Cell culture production is the core process in the scale-up in r-DNA technology: the proteins are produced by serial fermentation, in which the amount of cloning (number of functioning cells, volume of growth media, and protein production) are increased substantially in a stepwise fash-

ion. This cell culture process can employ a large fermenter wherein the cells float free in the medium, that is, a suspension culture. Alternatively, attachment technology, e.g., a fluidized bed fermenter or roller bottles, is used, wherein cells are fixed on a surface and are bathed in media.

The fermentation process requires careful control of the cell environment, including the nutrient medium, the air mixture (oxygen, nitrogen, carbon dioxide, and moisture), the breathability of the fermentation flasks (between flasks and ambient environment), the temperature of flasks, and the removal of cellular waste. Media shifts in roller bottles are performed periodically to maintain cell viability. Harvesting of protein from the medium is done at several times during the cell culture process, along with a step to concentrate the liquid mixture to create a crude multi-protein mixture. Throughout this fermentation period, the cells must be kept alive and functional without mutation, change, or contamination (e.g., viruses or oncogenes).

Purification of the crude protein is accomplished with a sequence of chromatographic steps to separate the desired protein from the crude mixture. The physical properties of the proteins (such as the size of molecules, the hydrophobic properties, and electrical charge) allow this separation to be done. The common chromatographic methods include gel-filtration, ion-exchange, affinity, reversed-phase, and hydrophobic techniques. Sterile filtration is then performed, to obtain a purified bulk of target protein without contaminating material or organisms. The final step, formulation, requires mixing the protein with a diluent plus stabilizers and buffers specific for that protein product. Filling and labeling of ready-to-use vials is necessary. Some biological products are lyophilized attempting to improve a product's shelf life, although reconstitution by practitioners then is required before use.

For microbial fermentation, the cell culture step uses one or several large fermenters containing the appropriate medium. Because of the high cell-growth rate, the fermenter has the capacity to force air into the media to maintain cell viability. At the end of fermentation, the cells can be killed without denaturing the proteins that are stored in intracellular inclusion bodies. The contents of the fermenter often are centrifuged to create a cell paste. Recovery of proteins from this microbial system usually calls for extraction of the inclusion bodies from the dead cells. The cells are broken up with a homogenizer or other high-shear device to obtain the inclusion bodies. A high-shear device also is employed to process the inclusion bodies and remove the bulk protein. The next step is purification and formulation similar to what we have described herein.

Quality control is a substantial challenge in the biotechnology industry, with >70 tests employed to ensure the integrity, quality, and activity of the intermediates as well as the final product. Biological products usually are proteins; as such, they are more complex compounds, have a more detailed and sophisticated manufacturing process, and are made in living cells, in contrast to the chemical synthesis and/or extraction processes used for drugs. These three features lead to a potentially greater opportunity for contamination, alteration, or denaturation of the protein product.

Four areas are addressed by quality control, that is, plasmids and host cells, bulk product, process validation, and final product batches. First, a guarantee of the integrity of host cells and plasmids necessitates, for example, chromosomal analysis of karyotype, contamination screens for viruses and oncogenes, and gene stability assessment. Second, testing of bulk product is extensive for a protein encompassing, for example, high-pressure liquid chromatographic analysis, radioimmunoassay, Western and Southern blot analyses, amino acid sequencing, peptide maps, and even bioassay. The goals of such a large number of sophisticated tests are to ensure structure, purity, potency, and activity of the protein. Third, final product is evaluated with the same tests as for bulk products, along with standard assessments of final formulation (such as tests of sterility, excipients, volume, appearance, and stability). Fourth, process validation tests are intended to address the efficiency and reliability of the manufacturing process.

TECHNIQUES OF BIOTECHNOLOGY

Although recombinant DNA technology is responsible for >90% of the available biological products, biotechnology has evolved to create several key technologies for product development, as represented in Table 2. Monoclonal antibody (MAb) technology has been in use for the past 20 years and has led to many diagnostic advances, e.g., ELISA (enzyme-linked immunosorbent assays) test kits for drug testing and OncoScint CR103, radiolabeled monoclonal antibodies for diagnosis of colorectal cancer. Therapeutic applications are limited at this

TABLE 2. *Techniques of biotechnology*

Recombinant DNA technology
Monoclonal antibody technique
Polymerase chain reaction
Gene therapy
Nucleotide blockade
Peptide technology
Carbohydrate technology

time to one product, antirejection (kidney) therapy with murononab, but many products are in development, especially for cancer therapy. Manufacture of a MAb requires creation of a hybridoma from mouse B cells, which produce specific antibodies against a target antigen, and from myeloma cells, which have long lives and produce large amounts of antibodies (7). The hybridoma possesses properties of both cell components. It is cloned and grown to yield antibodies, which are then harvested and purified.

The polymerase chain reaction is a biotechnology whereby substantial amplification (over several 100,000-fold) of a target nucleic acid sequence (e.g., a gene) is obtained (8). This enzymic reaction happens in repeated cycles of a three-step process: (a) DNA is denatured to separate the two strands; (b) a nucleic acid primer is hybridized to each DNA strand at a specific location in the nucleic acid sequence; and (c) DNA polymerase enzyme is added for extension of the primer along the DNA strand to copy the target nucleic acid sequence. Each cycle doubles the DNA molecules copied, and the cycle is repeated until sufficient DNA sequence material is copied, e.g., 20 cycles with a 90% success rate yields a 375,000 amplification of a DNA sequence.

Gene therapy is employed in inheritable diseases where a protein is either dysfunctional or not produced because of an abnormal gene (9). These common and life-threatening diseases include cystic fibrosis, hemophilia, sickle cell anemia, and diabetes. The gene responsible for the malady must be identified first, which was done, for example, for ADA (adenine deaminase) enzyme deficiency. This disease leads to a severely compromised immune deficiency (SCID), often causing death in childhood or adolescence. The normal human gene for the fully functional protein is cloned and inserted into a carrier, such as a harmless virus. A patient's cells, e.g., T lymphocytes, are grown in the laboratory, and the cells receive the gene from the viral carrier. The patient's cells start producing the missing protein to correct the deficiency. These cells with the extra functional gene then are returned to the patient, and the normal protein is produced and released, alleviating the disease. For SCID and other diseases, gene therapy is life-saving.

Nucleotide blockade or antisense focuses on arresting expression of dysfunctional messenger RNA or DNA (10). A complementary m-RNA is created to match an abnormal m-RNA. The two m-RNA strands complex together, preventing translation of the m-RNA to form disease-producing proteins. Anti-DNA strands also can be created to complex with DNA to form a triple helix. Oligonucleotides, or short single strands of nucleic acids, instead of the full m-RNA also can be employed to block RNA expression. Viral disease (herpes simplex and HIV) and cancer (oncogenes) are two targets of this form of biotechnology.

Peptide technology entails screening for polypeptide molecules that can mimic large proteins, thus affording more simple products that may be more stable and easier to produce. The peptides can serve as protein receptor agonists or antagonists. Carbohydrate chemistry is targeted at inflammatory and immune diseases involving leukocyte activation. Expression of carbohydrate molecules from cell surfaces (cell adhesion molecules) can be prevented, arresting cell migration or activation and, in turn, moderating the leukocyte-related components of inflammation.

These brief descriptions of biotechnology techniques now employed in product development expose the vast potential for many and varied biological products, which will modify or correct many diseases. The biological products now in development in clinical trials well exceed 100 molecules.

BIOLOGICAL PRODUCT AVAILABILITY

Fifteen biological molecules have been developed and marketed in the years from 1982 through mid-1993. Fourteen companies are marketing 20 products. Table 3 lists the molecules, products, companies, indications, and year of availability. Some products are used for single indications [such as insulin for diabetes (1982) or Factor VIII for hemophilia (1992)]. However, many products are approved first for one major indication and then, after substantial clinical research into additional indications, are approved for other uses at a later time. Examples include interferon α-2b (Intron A) for hairy cell leukemia (1986), AIDS-related Kaposi's sarcoma (1988), hepatitis C (1991), and hepatitis B (1992), and epoetin α (EPOGEN, Procrit) for ane-

TABLE 3. *Biological products*[a]

Generic name	Trade name	Therapeutic area
Human insulin	Humulin (Lilly)	Insulin-dependent diabetes mellitus (1982)
Human growth hormone	Protropin (Genentech) Humatrope (Lilly)	Human growth hormone deficiency in children (1985)
Hepatitis B vaccine	Engerix-B (SmithKline Beecham) Recombivax HB (MSD)	Hepatitis B prophylaxis (1986)
Interferon α-2a	Roferon A (Hoffman LaRoche)	Hairy cell leukemia (1986); AIDS-related Kaposi's sarcoma (1988)
Interferon α-2b	Intron A (Schering-Plough)	Hairy cell leukemia (1986); AIDS-related Kaposi's sarcoma (1988); chronic hepatitis types B (1992) and C (non-A, non-B) (1991); condylomata acuminata (1988)
Muromonab-CD3	Orthoclone OKT 3 (Ortho Biotech)	Acute allograft rejection in renal transplant patients (1986)
Alteplase	Activase (Genentech)	Acute myocardial infarction (1987); pulmonary embolism (1990)
Epoetin α	Epogen (Amgen) Procrit (Ortho Biotech)	Certain anemias—chronic renal disease (1989); AIDS (1991); cancer chemotherapy (1993)
Interferon α-n3	Alferon N (Interferon Sciences)	Condylomata acuminata (1989)
Interferon γ-1b	Actimmune (Genentech)	Chronic granulomatous disease (1990)
Filgrastim (G-CSF)	Neupogen (Amgen)	Febrile neutropenia (infection) due to myelosuppressive chemotherapy (1991)
Sargramostim (GM-CSF)	Leukine (Immunex) Prokine (Heochst-Roussel)	Myeloid reconstitution after bone marrow transplantation (1991)
Aldesieukin	Proleukin (Chiron)	Metastatic renal cell carcinoma (1992)
Factor VIII	KoGenate (Miles) Recombinate (Baxter)	Hemophilia A (1992)
Interferon β	Betaseron (Chiron, Berlex)	Multiple sclerosis (1993)

G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte–macrophage colony-stimulating factor.

[a] List of products is given in order of date of first product approval.

mia stemming from chronic renal disease (1989), from AIDS-related antiviral therapy (1991), and from cancer chemotherapy (1993). The disease entities treated by these 15 molecules are commonplace and address most organ systems and 18 separate conditions. They can be used in endocrinology (diabetes, growth), hepatitis (types B and C), oncology (hairy cell leukemia, AIDS-related Kaposi's sarcoma, renal cell carcinoma, neutropenia, and infections from chemotherapy), anemias, cardiology (heart attacks, pulmonary embolism, hemophilia), neurology (multiple sclerosis), and organ (kidney) rejection. This brief description establishes the profound impact of biotechnology on health care, bringing therapeutic agents to treat diseases that often were previously untreatable. A new arm of therapy, biotherapy, has been created.

FUTURE BIOTECHNOLOGY PRODUCTS

The first decade of biotechnology drug development produced an admirable record of one new biological molecule per year, on average. This early success is expected to be followed by a host of products in similar and new therapeutic areas. According to 1991 Pharmaceutical Manufacturing Association (PMA) data, 132 biological products were in clinical development and 21 molecules were at the Federal Drug Administration (FDA) for regulatory review and potential approval (11). The time frame to bring such new biological products to market (patient availability) has been documented to average ~4.8 years in the United States and 3 years in Europe; two-thirds of this time is consumed by clinical research and one-third by regulatory review

(12). According to PMA data, from 1982 to 1992, FDA review of product license applications for biological products averaged 24 months, and new indications took 30 months for review and approval. Over the past 5 years, the research pipeline has grown from <80 to >130 products (60% growth), and the expected time required for product research and approval is ~5 years.

These new biotherapies incorporate both extensions of current product categories and whole new entities. Table 4 cites 10 product categories for future biological products and includes >150 new products or indications that are in development (13). Most therapeutic areas are being targeted by biotechnology, but oncology has the largest number of products; some two-thirds to three-quarters of interferons, interleukins, and monoclonal antibodies are used in this area. In vaccines, AIDS is the major focus. Tissue-growth factors, including epidermal growth factor, fibroblast growth factor, and platelet-derived growth factor, address repair of wounds due to injury, surgery, chronic disease, or decubiti. Clotting factors include proteins to improve clotting and to dissolve clots. The last category (labeled "various") includes such molecules as dismutases, tumor necrosis factor, and soluble CD4.

TABLE 4. *Future biotherapies*

Category	Number of indications and products
Antisense	5
Clotting factors	8
Hematopoietic factors	12
Hormones	7
Interferons	16
Interleukins	13
Monoclonal antibodies	58
Tissue growth factors	11
Vaccines	18
Various	>15

See Burrill and Lee for summary (13).

Biological products have become the mainstay of therapy for major diseases already benefiting millions of patients each year, with only 15 molecules and 20 products in use to date. Future products and indications exceed 100 in number. The full impact of these therapies and the benefits to patients still are being discovered. Biotherapy holds the promise of a natural therapy, wherein proteins, nucleotides, and other human molecules are recreated and employed to ameliorate or even cure disease. We are on the threshold of a pharmacologic and therapeutic revolution.

REFERENCES

1. Drews J. Into the 21st century: biotechnology and the pharmaceutical industry in the next ten years. *Bio/technology* 1993;11:S16–S20.
2. Dibner MD. Tracking trends in the U.S. biotechnology. *Biol technology* 1991;9:1334–7.
3. Miller L, Escrow J, Malloy M, eds. *Biotechnology industry evaluation: 1992 sector analysis*. New York: Paine Webber Inc., Health Care Group, 1992.
4. Klausner A. Back to the future: biotech product sales 1983–1993. *Bio/technology* 1993;11:S35–6.
5. Van Noordwijk J. Quality assurance of products manufactured by recombinant DNA technology: introduction and elements of a philosophy. *Arzneimittalforschung* 1988;38: 943–7.
6. Sattele DB, ed. *Biotechnology in perspective*. London: Hobsons Publishing, 1990.
7. Pirofski L, Gasadevall A, Rodriguez L, Zuckier LS, Scharff MD. Current state of hybridoma technology. *J Clin Immunol* 1990;10(suppl. 6):5s–14s.
8. Lipner M. The polymerase chain reaction: amplifying its role in research and beyond. *Oncol Times* 1992;14–5.
9. Goldspiel BR, Green L, Calis KA. Human gene therapy. *Clin Pharmacy* 1993;12:488–505.
10. Nagel KM, Holstad SG, Isenberg KE. Oligonucleotide pharmacotherapy: an antigene strategy. *Pharmacotherapy* 1993; 13:177–88.
11. In development: biotechnology medicines. Washington, D.C.: Pharmaceutical Manufacturing Association, 1991.
12. Bienz-Tadmor B. Biopharmaceuticals go to market: patterns of worldwide development. *Bio/technology* 1993;11:168–72.
13. Burrill GS, Lee Jr KB. *Biotech '92: promise to reality*. An industry annual report. San Francisco: Ernst and Young, 1991.

STIC-ILL

From: Huynh, Phuong N.
Sent: Tuesday, September 21, 2004 4:59 PM
To: STIC-ILL
Subject: RE: 09/494,096 rush

Importance: High

Please deliver:

J Allergy Clin. Immunol. 109:143-149, 2002.
J. Allergy Clin. Immunol. 105:S 184 (Abstract 555), 2000.
J. Allergy Clin. Immunol. 105:378-384, 2000.
J. Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
J Allergy Clin. Immunol. 105:S140 (Abstract 423), 2000
Ann. NY Acad Sci. 964:69-85, 2002.

Ferreira *et al*, Adv Exp Med Biol 409: 127-135, 1996.

Fasler *et al*, J. Allergy and Clinical Immunology 101(4 pt 1): 521-30, April 1998.

Burks *et al*, Eur. J. Biochem. 245: 334-339, April 1997.

Stanley *et al*, Archives of Biochemistry and Biophysics 342(2): 244-253, June 1997.

Skolnick *et al*, Trends in Biotech. 18(1): 34-39, Jan 2000.

Colman *et al*, A structural view of immune recognition by antibodies, pages 33-36, 1994.

Evens *et al*, Therapeutic Drug Monitoring 15: 514-520, 1993.

Thanks,
Neon
Art unit 1644
Room REM 3D75
Tel (571) 272-0846

Eur. J. Biochem. 245, 334–339 (1997)

Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity

A. WESLEY BURKS[1], David SHIN[2], Gael COCKRELL[1], J. Steven STANLEY[1,2], Ricki M. HELM[1] and Gary A. BANNON[2]

[1] Department of Pediatrics, University of Arkansas for Medical Sciences, Arkansas Children's Hospital Research Institute, Little Rock, Arkansas, USA

[2] Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Arkansas Children's Hospital Research Institute, Little Rock, Arkansas, USA

(Received 9 January 1997) – EJB 97 0031/1

Peanut allergy is a significant health problem because of the prevelance and potential severity of the allergic reaction. Serum IgE from patients with documented peanut hypersensitivity reactions and overlapping peptides were used to identify the IgE-binding epitopes on the major peanut allergen, Ara h 1. At least twenty-three different linear IgE-binding epitopes, located throughout the length of the Ara h 1 protein, were identified. All of the epitopes were 6–10 amino acids in length, but there was no obvious sequence motif shared by all peptides. Four of the peptides appeared to be immunodominant IgE-binding epitopes in that they were recognized by serum from more than 80% of the patients tested and bound more IgE than any of the other Ara h 1 epitopes. Mutational analysis of the immunodominant epitopes revealed that single amino acid changes within these peptides had dramatic effects on IgE-binding characteristics. The identification and determination of the IgE-binding capabilities of core amino acids in epitopes on the Ara h 1 protein will make it possible to address the pathophysiologic and immunologic mechanisms regarding peanut hypersensitivity reactions specifically and food hypersensitivity in general.

Keywords: food allergy; IgE-binding epitope; epitope analysis.

Approximately 8% of children and 1–2% of adults have some type of food allergy [1]. Peanuts, fish, tree nuts, and shellfish account for the majority of food hypersensitivity reactions in adults, while peanuts, milk, and eggs cause over 80% of food hypersensitivity reactions in children [2].

Food hypersensitivity reactions occur shortly after contact of a specific allergen with its corresponding IgE antibodies which are bound to mast cells. Allergen-specific IgE, when cross-linked by the respective allergen, activates the mast cells to release histamine, heparin, and other mediators responsible for the clinical symptoms observed. Thus the IgE-binding epitopes of the allergens play an important role in the disease process. Their characterization will provide a better understanding of the human immune response involved in food hypersensitivity reactions. If improved diagnostic and therapeutic capabilities are to be developed it is important to determine the primary structure and frequency of recognition of any IgE-binding epitopes contained within the allergen.

Various studies have shown that the most allergenic portion of the peanut is the protein fraction of the cotyledon [3]. A major allergen found in the cotyledon is the peanut protein, Ara h 1 [4]. The Ara h 1 allergen belongs to the vicilin family of seed storage proteins [5]. This protein is recognized by greater than 90% of peanut-sensitive patients, thus establishing it as an important allergen [4]. The majority of serum IgE recognition of the Ara h 1 allergen appears to be due to epitopes within this protein that are linear amino acid sequences that do not contain significant amounts of carbohydrate [4, 5]. Previous results have demonstrated similarity between the level of IgE binding to recombinant Ara h 1 protein and the native form of this allergen when individual patient serum was tested [5]. These results indicated that the recombinant protein could be considered for use in both diagnostic and immunotherapeutic approaches to peanut hypersensitivity.

Because of the prevalence and severity of peanut hypersensitivity reactions in both children and adults, and the role B-cell epitopes play in these reactions, we set out to map and characterize the major IgE epitopes of the Ara h 1 allergen. In this study, we report the primary structure of the Ara h 1 IgE-binding epitopes recognized by peanut hypersensitive individuals. Four epitopes that bound peanut-specific serum IgE from more than 80% of patients tested were identified. The amino acids important to peanut-specific IgE recognition of these epitopes were then determined for the purpose of using them in future diagnostic and immunotherapeutic approaches to this disease.

MATERIALS AND METHODS

Serum IgE. Serum from 15 patients with documented peanut hypersensitivity reactions (mean age, 25 y) was used to identify the Ara h 1 IgE-binding epitopes. Each of these individuals had a positive immediate prick skin test to peanut and either a positive double-blind placebo-controlled food challenge or a convincing history of peanut anaphylaxis (laryngeal edema, severe wheezing, and/or hypotension). Representative individuals with elevated serum IgE levels (who did not have peanut-specific IgE or peanut hypersensitivity) were used as controls in these studies. In some instances, a serum pool was made by mixing equal aliquots of serum IgE from each of the 15 patients with peanut hypersensitivity. This pool was then used in immu-

Correspondence to G. A. Bannon, University of Arkansas for Medical Sciences, Slot 516, 4301 W. Markham, Little Rock, AR 72205, USA

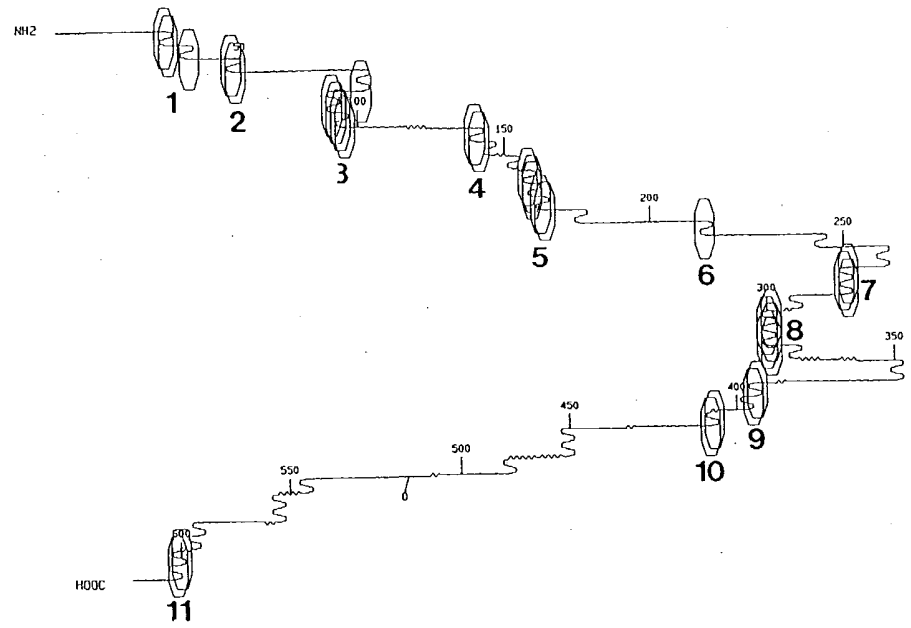


Fig. 1. There are multiple predicted antigenic sites on the Ara h 1 allergen. The amino acid sequence of the Ara h 1 protein was analyzed for potential antigenic sites by the Jameson and Wolf (1988) algorithm. There were 11 (1–11) predicted regions that contained multiple antigenic sites (octagons) along the entire length of the molecule. The amino acid sequence is represented as α-helical (sinusoidal curve), β-sheet (saw tooth curve), and coil (flat sinusoidal curve) conformations. β turns are denoted by chain reversals.

noblot analysis experiments to determine the IgE-binding characteristics of the population. At least 5 ml venous blood was drawn from each patient and allowed to clot, and the serum collected. All studies were approved by the Human Use Advisory Committee at the University of Arkansas for Medical Sciences.

Computer analysis of Ara h 1 sequence. Analysis of the Ara h 1 gene [5] and peptide sequences was performed on the University of Arkansas for Medical Sciences' Vax computer using the Wisconsin DNA analysis software package. The predicted antigenic regions on the Ara h 1 protein are based on algorithms developed by Jameson and Wolf [6] that relate antigenicity to hydro-philicity, secondary structure, flexibility, and surface probability.

Peptide synthesis. Individual peptides were synthesized on a cellulose membrane containing free hydroxyl groups using Fmoc-amino acids according to the manufacturer's instructions (Genosys Biotechnologies). Synthesis of each peptide was started by esterification of an Fmoc-amino acid to the cellulose membrane. After washing, all residual amino functions on the sheet were blocked by acetylation to render them unreactive during the subsequent steps. Fmoc protective groups were then removed by addition of piperidine to render nascent peptides reactive. Each additional Fmoc-amino acid is esterified to the previous one by this same process. After addition of the last amino acid in the peptide, the amino acid side chains were de-protected using a mixture of 1:1:0.05 (by vol.) dichloromethane/trifluoroacetic acid/triisobutylsilane, followed by washing with dichloromethane and methanol. Membranes containing synthesized peptides were either probed immedately with serum IgE or stored at −20°C until needed.

IgE-binding assay. Cellulose membranes containing synthesized peptides were incubated with the serum pool or individual serum from patients with peanut hypersensitivity diluted (1:5) in a solution containing Tris/NaCl (10 mM Tris/HCl, 500 mM NaCl, pH 7.5) and 1% bovine serum albumin for at least 12 h at 4°C or 2 h at room temperature. The primary antibody was detected with ^{125}I-labeled anti-IgE antibody (Sanofi Pasteur Diagnostics).

RESULTS

There are multiple IgE-binding regions throughout the Ara h 1 protein. The Ara h 1 amino acid sequence was analyzed for potential antigenic epitopes using computer-based algorithims. There were 11 possible antigenic regions, each containing multiple antigenic sites, predicted by this analysis along the entire length of the molecule (Fig. 1).

77 overlapping peptides representing the entire length of the Ara h 1 protein were synthesized to determine if the predicted antigenic regions, or any other regions, were recognized by serum IgE. Each peptide was 15 amino acids long and offset from the previous peptide by eight amino acids. In this manner, the entire length of the Ara h 1 protein could be studied in large overlapping fragments. These peptides were then probed with a pool of serum IgE from 15 patients with documented peanut hypersensitivity or with serum IgE from a representative control patient with no food allergy. Serum IgE from the control patients did not recognize any of the synthesized peptides. In contrast, there are 12 IgE-binding regions (D1–D12) along the entire length of the Ara h 1 protein recognized by IgE from this population of patients with peanut hypersensitivity (Fig. 2). These IgE-binding regions represent amino acid residues 35–72, 89–112, 121–176, 289–326, 337–350, 361–374, 393–416, 457–471, 489–513, 521–535, 544–583, and 593–607. In general, the predicted antigenic regions (Fig. 2, boxed areas P1–P11) contained or were part of those that were determined (Fig. 2, shaded areas D1–D12) by actual IgE binding. However, there were two predicted antigenic regions (amino acids A221–230; amino acids 263–278) that were not recognized by serum IgE from peanut hypersensitive individuals. In addition, there were numerous IgE-binding regions found in the Ara h 1 protein between amino acids 450–600 (Fig. 2).

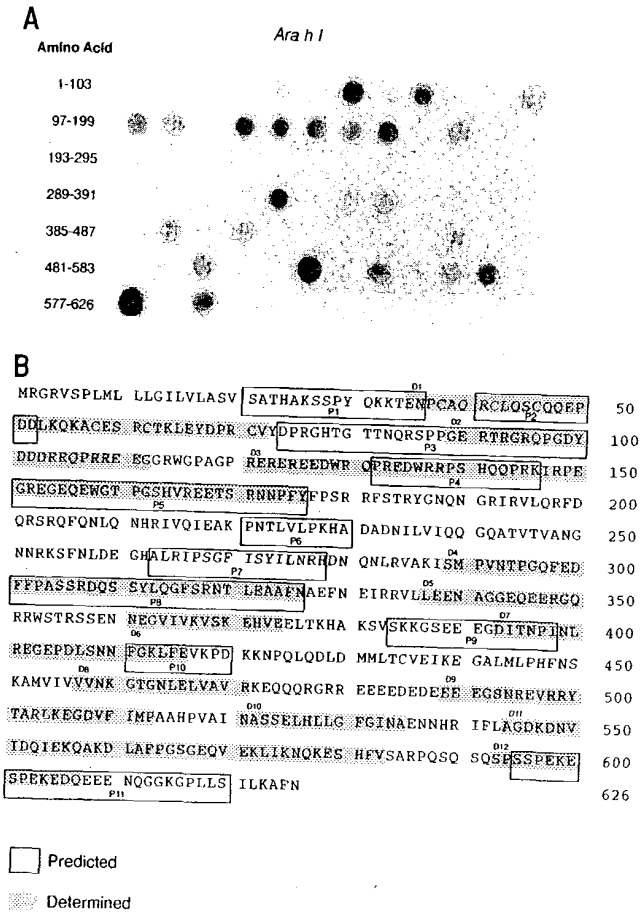


Fig. 2. Multiple IgE-binding regions identified on the Ara h 1 allergen. (A) Epitope mapping was performed on the Ara h 1 allergen by synthesizing the entire protein in 15-amino-acid-long overlapping peptides that were offset from each other by eight amino acids. These peptides were then probed with a pool of serum IgE from 15 patients with documented peanut hypersensitivity. The position of the peptides within the Ara h 1 protein are indicated on the left-hand side of A. (B) The amino acid sequence of the Ara h 1 protein is shown. The numbered boxes correspond to the predicted antigenic regions (P1–P11). The hatched areas (D1–D12) correspond to the IgE-binding regions shown in the upper panel.

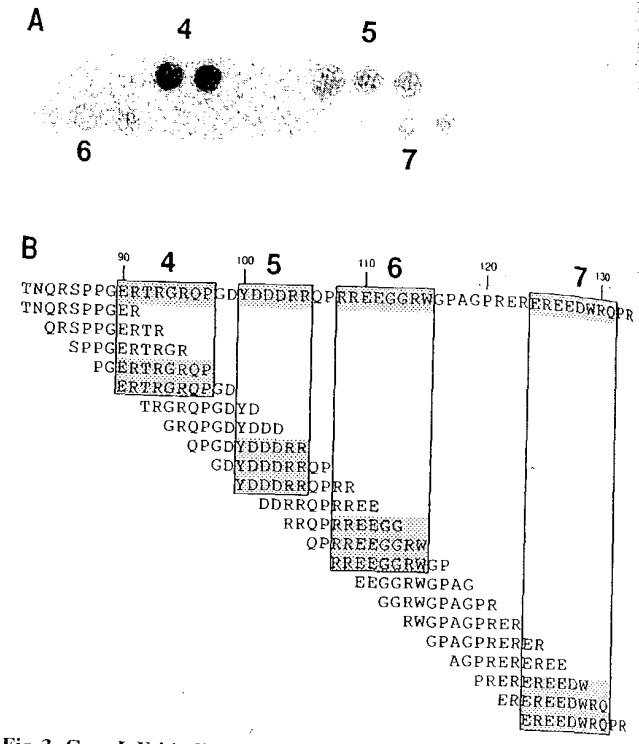


Fig. 3. Core IgE-binding epitopes identified on the Ara h 1 allergen. (A) Detailed epitope mapping was performed on IgE-binding regions identified in Fig. 2 by synthesizing ten-amino-acid-long peptides offset from each other by two amino acids. These peptides were then probed with a pool of serum IgE from 15 patients with documented peanut hypersensitivity. The data shown represents regions D2 and a portion of D3 encompassing amino acid residues 82–133. Numbers correspond to peptides as shown in Table 1. (B) The amino acid sequence (residues 82–133) of Ara h 1 that was tested in A is shown. Shaded areas of boxes correspond to common IgE-binding amino acids of the spots shown in A.

To determine the amino acid sequence of the IgE-binding sites, small overlapping peptides spanning each of the larger IgE-binding regions identified in Fig. 2 were synthesized. By synthesizing smaller peptides (10 amino acids long) that were offset from each other by only two amino acids, it was possible to identify individual IgE-binding epitopes within the larger IgE-binding regions of the Ara h 1 molecule. A representative immunoblot and the respective amino acid sequence of the binding region D2–D3 (amino acids 82–133) is shown in Fig. 3. Four epitopes (Fig. 3, 4–7) were identified in this region. Similar blots were completed for the remaining IgE-binding regions to identify the core amino acid sequences for each IgE epitope. Table 1 summarizes the 23 IgE-binding epitopes (peptides 1–23) and their respective positions in the Ara h 1 molecule. The most common amino acids found were acidic (D, E) and basic (K, R) residues comprising 40% of all amino acids found in the epitopes. There were no obvious amino acid sequence motifs shared by all the IgE-binding epitopes.

Identification of common Ara h 1 epitopes recognized by serum IgE from patients with peanut hypersensitivity. In an

Table 1. Ara h 1 IgE-binding epitopes. The underlined portions of each peptide are the smallest IgE-binding sequences as determined by the analysis described in Fig. 3.

Peptide	Amino acid sequence	Ara h 1 positions
1	AKSSPYQKKT	25–34
2	QEPDDLKQKA	48–57
3	LEYDPRLVYD	65–74
4	GERTRGRQPG	89–98
5	PGDYDDDRRQ	97–105
6	PRREEGGRWG	107–116
7	RERREEDWRQP	123–132
8	EDWRRPSHQQ	134–143
9	QPRKIRPEGR	143–152
10	TPGQFEDFFP	294–303
11	SYLQEFSRNT	311–320
12	FNAEFNEIRR	325–334
13	EQEERGQRRW	344–353
14	DITNPINLRE	393–402
15	NNFGKLFEVK	409–418
16	GTGNLELVAV	461–470
17	RRYTARLKEG	498–507
18	ELHLLGFGIN	525–534
19	HRIFLAGDKD	539–548
20	IDQIEKQAKD	551–560
21	KDLAFPGSGE	559–568
22	KESHFVSARP	578–587
23	PEKESPEKED	597–606

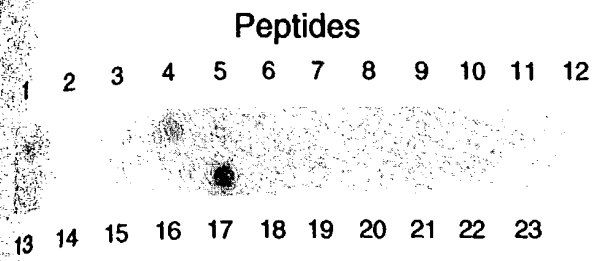


Fig. 4. Ara h 1 epitopes recognized by serum IgE from a patient with peanut hypersensitivity. IgE-binding peptides (1–23) were synthesized as ten-amino-acid peptides and probed with serum IgE from a patient with documented peanut hypersensitivity.

effort to determine which, if any, of the 23 epitopes was immunodominant, each set of 23 peptides was probed individually with serum IgE from ten different patients. An epitope can be considered immunodominant if it is recognized by serum IgE from the majority of patients with peanut hypersensitivity or if the serum IgE that recognizes a peptide represents the majority of Ara h 1-specific IgE found in a patient. Serum from five individuals randomly selected from the 15 patient serum pool and an additional five sera from peanut-hypersensitive patients not represented in the serum pool were used to identify the commonly recognized epitopes. An immunoblot strip containing these peptides that has been incubated with an individual patient's serum is shown in Fig. 4. This patient's serum IgE recog-

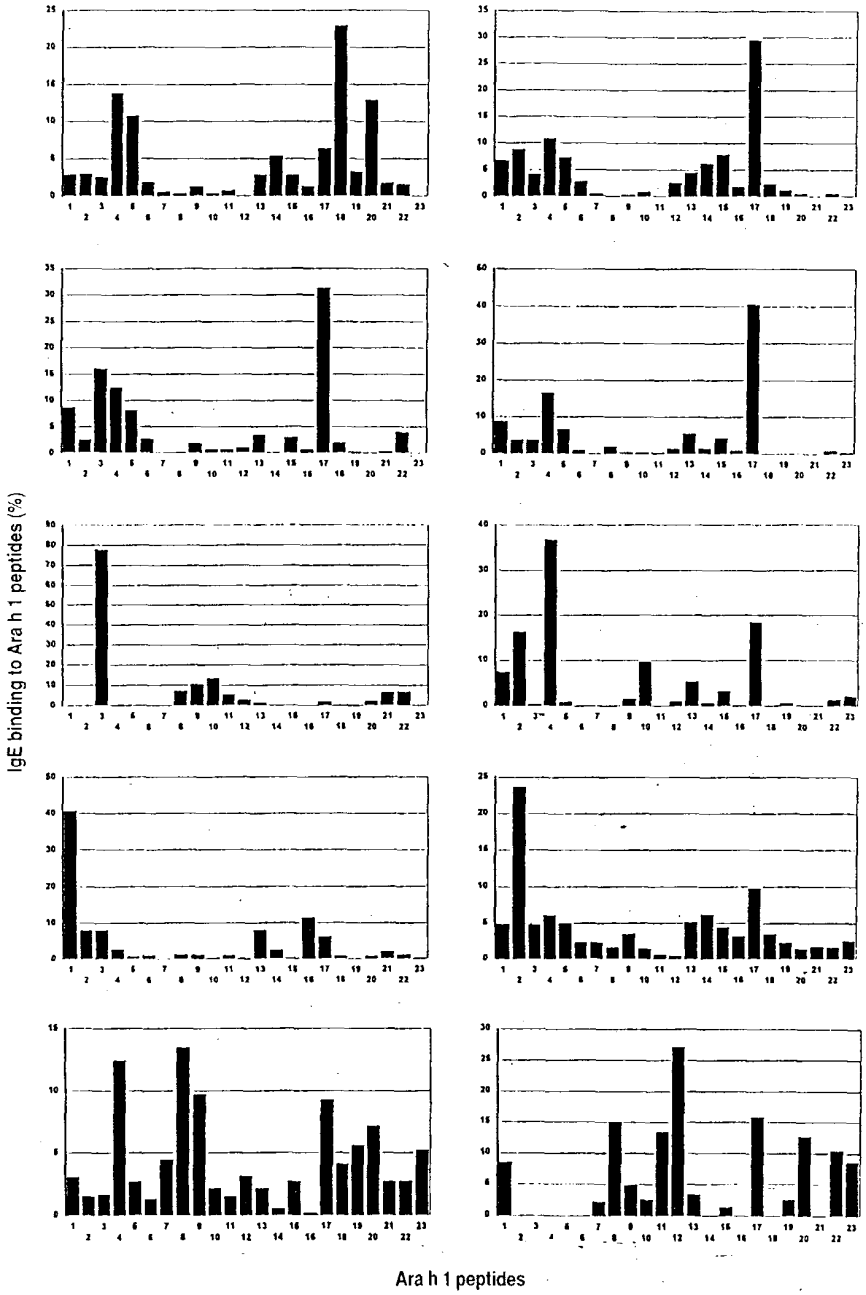


Fig. 5. Identification of the immunodominant Ara h 1 epitopes. IgE-binding peptides (1–23) were synthesized as described in Fig. 4 and probed with serum IgE from 10 patients (A–J) with peanut hypersensitivity. Serum from five individuals randomly selected from the 15-patient serum pool and an additional five sera from peanut-hypersensitive patients not represented in the serum pool were used to identify the immunodominant epitopes. The relative intensity of IgE binding to each peptide was then determined by densitometry and expressed as a percentage of the patient's total IgE binding to all of the Ara h 1 peptides.

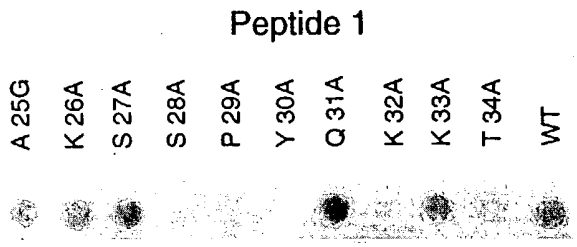


Fig. 6. Single amino acid changes to peptide 17 result in loss of IgE binding to this epitope. Peptide 17 was synthesized with an alanine residue substituted for one of the amino acids in this peptide and probed with a pool of serum IgE from 15 patients with documented peanut hypersensitivity. The letters across the top of each panel indicate the one-letter amino acid code for the residue normally at that position and the amino acid that was substituted for this residue. The numbers indicate the position of each residue in the Ara h 1 protein. WT, indicates the wild-type peptide (no amino acid substitutions).

nized peptides 1, 4, 5, 8, 13, 15, and 17. Peptides 2, 3, 6, and 9 were detected but could only be minor epitopes considering the amount of IgE that they bound. The remaining patients' serum IgE were tested in the same manner and the intensity of IgE binding to each spot was determined as a function of that patient's total IgE binding to these 23 epitopes (Fig. 5). All of the patient sera tested (10/10) recognized multiple peptides. The most commonly recognized peptides were those that contained epitopes 1, 3, 4, 13, 17, and 22. These epitopes were recognized by IgE from at least 80% of the patient sera tested (8/10). In addition, epitopes 1–4, 8, 12, and 17, when recognized, bound more serum IgE from individual patients than any of the other epitopes. These results indicate that peptides 1, 3, 4, and 17 contain the immunodominant epitopes of the Ara h 1 protein.

IgE-binding characteristics of the immunodominant Ara h 1 epitopes. The amino acids essential to IgE binding in epitopes 1, 3, 4, and 17 were determined by synthesizing duplicate peptides with single amino acid changes at each position. These peptides were then probed with pooled serum IgE from 15 patients with peanut hypersensitivity to determine if the change affected peanut-specific IgE binding. An immunoblot strip containing the wild-type and mutated peptides of epitope 1 is shown in Fig. 6. The pooled serum IgE did not recognize this peptide, or binding was drastically reduced, when alanine was substituted for each of the amino acids at positions 28–30, or 32. In contrast, the substitution of an alanine for glutamine residue at position 31 resulted in increased IgE binding. The remaining immunodominant Ara h 1 epitopes were tested in the same manner and the intensity of IgE binding to each spot was determined as a percentage of IgE binding to the wild-type peptide (Fig. 7). In general, each epitope could be mutated to a non-IgE-binding peptide by the substitution of an alanine for a single amino acid residue. There was no obvious position within each peptide that, when mutated, would result in loss of IgE binding. Furthermore, there was no consensus in the type of amino acid that, when changed to alanine, would lead to loss of IgE binding.

DISCUSSION

The development of an IgE response to an allergen involves a series of interactions between T cells and B cells. B cells bearing appropriate antigen-specific surface immunoglobulins interact with proliferating allergen-specific T-cells which leads to isotype switching and the generation of antigen-specific IgE. The antigen-specific IgE then binds to the FcεRI receptors of mast cells and basophils. This establishes a condition in which the IgE-bound mast cells can trigger a potentially life threaten-

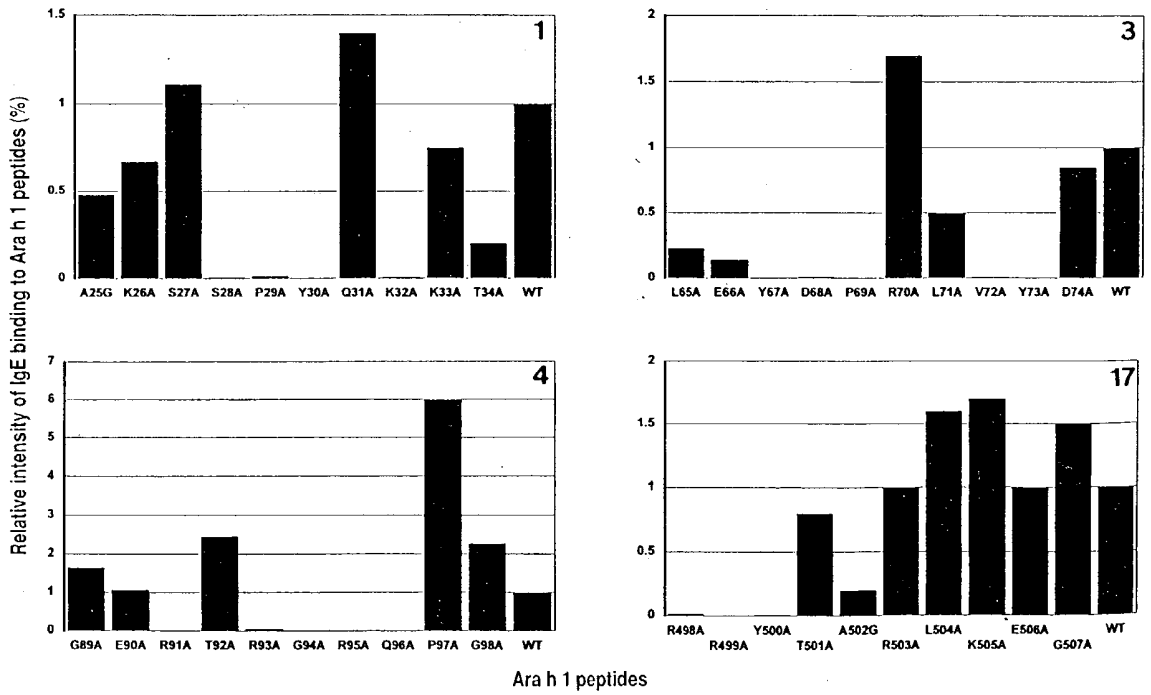


Fig. 7. All of the immunodominant Ara h 1 peptides can be mutated to non-IgE binding epitopes by single amino acid changes. All of the immunodominant Ara h 1 epitopes (peptides 1, 3, 4, and 17) were synthesized with an alanine or glycine residue substituted for one of the amino acids and probed with a pool of serum IgE from 15 patients with documented peanut hypersensitivity. The relative intensity of IgE binding to each peptide was then determined by densitometry and expressed as a percentage of IgE binding to the non-mutated peptide (WT). The letters across the bottom of each panel indicate the one-letter amino acid code for the residue normally at that position and the amino acid that was substituted for this residue. The numbers indicate the position of each residue in the Ara h 1 protein.

ing immediate hypersensitive reaction when next exposed to the allergen. Because antigen-specific IgE plays such a critical role in the etiology of allergic disease, determination of allergen-specific, IgE-binding epitopes is an important first step toward a better understanding of this complex disease process.

In the present study, we have determined that there were multiple antigenic sites predicted for the Ara h 1 allergen. In general, as has been found with other allergens [7, 8], there was good agreement between those residues predicted by computer analysis and B-cell epitopes determined by experimental analysis of overlapping peptides. This correlation between predicted and determined epitopes is probably due to the ability of the computer model to predict which regions of the molecule are exposed on the surface of the allergen, making them accessible to immunoglobulin interactions. There are at least 23 different IgE recognition sites on the major peanut allergen Ara h 1. These sites are distributed throughout the protein. The identification of multiple epitopes on a single allergen is not novel. Allergens from cow milk [9], codfish [10], hazel [11], soy [12], and shrimp [13] have all been shown to contain multiple IgE-binding epitopes. The observation that most of these proteins have multiple IgE-binding sites probably reflects the polyclonal nature of the immune response to them and may be a necessary step in establishing a protein as an allergen.

The elucidation of the major IgE-binding epitopes on Ara h 1 may also enable us to better understand the immunopathogenic mechanisms involved in peanut hypersensitivity. Recent evidence suggests that there is a preferential variable heavy-chain usage in IgE synthesis and a direct switching from IgM production to IgE synthesis [14]. This would suggest that epitopes responsible for antigen-specific IgE antibody production may differ from those promoting antigen-specific IgG antibodies. Immunotherapeutic approaches utilizing peptides representing IgG epitopes may be able to shift the balance of antigen-specific antibody production from IgE to IgG. We are currently identifying which of the IgE-binding epitopes also bind IgG to determine if this would be a feasible strategy for patients with peanut hypersensitivity.

Four of the Ara h 1 peptides appear to be immunodominant IgE-binding epitopes in that they are recognized by more than 80% of patient sera tested. Interestingly, epitope 17, which is located in the C-terminal end of the protein (amino acids 498–507), is in a region that shares significant sequence similarity with vicilins from other legumes [15]. The amino acids important for IgE binding also appear to be conserved in this region and may explain the possible cross-reacting antibodies to other legumes that can be found in sera of patients with a positive double-blind placebo-controlled food challenge to peanuts. Epitopes 1, 3, and 4 located in the N-terminal portion (amino acids 25–34, 65–74, and 89–98) of the protein, appear to be unique to this peanut vicilin and do not share any significant sequence similarity with vicilins from other legumes [15]. In addition, the amino acids important to IgE binding in this region are not conserved. These findings may enable us to develop more sensitive and specific diagnostic tools and lead to the design of novel therapeutic agents to modify the allergic response to peanuts.

Our data show that it may be possible to mutate the Ara h 1 allergen to a protein that no longer binds IgE. This raises the possibility that an altered Ara h 1 gene could be used to replace its allergenic homologue in the peanut genome. In this manner, a hypoallergenic peanut could be developed that may blunt allergic reactions in sensitive individuals who inadvertently ingest this food. Furthermore, a hypoallergeic peanut may prevent the development of peanut hypersensitivity in individuals genetically predisposed to this food allergy. Because the Ara h 1 gene product is such an abundant and integral seed storage protein [15], a prudent approach would require that the altered vicilin retain as much of its native function and properties as possbile. Given the potential severity of peanut allergic reactions and the widespread use of peanuts in consumer foods, this possibility is actively being pursued in our laboratories.

This work was supported in part by grants from the National Institutes of Health (AI33596) and the Clarissa Sosin Research Foundation.

REFERENCES

1. Jansen, J. J., Kardinaal, A. F. M., Huijber, G., Vleig-Boerstra, B. J., Martens, B. P. & Ockhuizen, T. (1994) Prevelance of food allergy and intolerance in the adult Dutch population, *J. Allergy Clin. Immunol. 93*, 446–456.
2. Sampson, H. A. (1988) The role of food alergy and mediator release in atopic dermatitus, *J. Allergy Clin. Immunol. 81*, 635–645.
3. Taylor, S. L., Busse, W. W., Sachs, M. I., Parker, J. L. & Yunginger, J. W. (1981) Peanut oil is not allergenic to peanut sensitive individuals, *J. Allergy Clin. Immunol. 68*, 372–375.
4. Burks, A. W., Williams, L. W., Helm, R. M., Connaughton, C., Cockrell, G. & O'Brien, T. (1991) Identification of a major peanut allergen *Ara h* I, in patients with atopic dermatitus and positive peanut challenge, *J. Allergy Clin. Immunol. 88*, 172–179.
5. Burks, A. W., Cockrell, G., Stanley, J. S., Helm, R. M. & Bannon, G. A. (1995) Recombinant peanut allergen *Ara h* I expression and IgE binding in patients with peanut hypersensitivity, *J. Clin. Invest. 96*, 1715–1721.
6. Jameson, B. A. & Wolf, J. (1988) The antigenic index: a novel algorithm for predicting antigenic determinants, *Comput. Appl. Biosci. 4*, 181–186.
7. van Kampen, V., Becker, W. M., Chen, Z., Rihs, H. P., Mazur, G., Raulf, M., Liebers, V., Isringhausen-Bley, S. & Baur, X. (1994) Analysis of B-cell epitopes in the N-terminal region of Chi t I component III using monoclonal antibodies, *Mol. Immunol. 31*, 1133–1140.
8. Breiteneder, H., Ferreira, F., Reikerstorfer, A., Duchene, M., Valenta, R., Hoffman-Sommergruber, K., Ebner, C., Breitenbach, M., Kraft, D. & Scheiner, O. (1992) Complementary DNA cloning and expression in *Escherichia coli* of Aln g I, the major allergen in pollen of alder (Alnus glutinosa), *J. Allergy Clin. Immunol. 90*, 909–917.
9. Ball, G., Shelton, M. J., Walsh, B. J., Hill, J., Hosking, C. S. & Howden, M. E. (1994) A major continuous allergenic epitope of bovine β-lactoglobulin recognized by human IgE binding, *Clin. Exp. Allergy 24*, 758–764.
10. Aas, K. & Elsayed, S. (1975) Physico-chemical properties and specific activity of a purified allergen (codfish), *Dev. Biol. Stand. 29*, 90–98.
11. Elsayed, S., Holen, E. & Dybendal, T. (1989) Synthetic allergenic epitopes from the amino-terminal regions of the major allergens of hazel and birch pollen, *Int. Arch. Allergy Appl. Immunol. 89*, 410–415.
12. Herian, A. M., Taylor, S. L. & Bush, R. K. (1990) Identification of soybean allergens by immunoblotting with sera from soy-allergic adults, *Int. Arch. Allergy Appl. Immunol. 92*, 193–198.
13. Shanti, K. N., Martin, B. M., Nagpal, S., Metcalf, D. D. & Rao, P. V. (1993) Identification of tropomyosin as the major shrimp allergen and characterization of its IgE-binding epitopes, *J. Immunol. 151*, 5354–5363.
14. van der Stoep, N., Korver, W. & Logtenberg, T. (1994) *In vivo* and *in vitro* IgE isotype switching in human B lymphocytes: evidence for a predominantly direct IgM to IgE class switch program, *Eur. J. Immunol. 24*, 1307–1311.
15. Gibbs, P. E., Strongin, K. B. & McPherson, A. (1989) Evolution of legume seed storage proteins – a domain common to legumins and vicilins is duplicated in vicilins, *Mol. Biol. Evol. 6*, 614–623.

Evidence for an early appearance of modern post-switch isotypes in mammalian evolution; cloning of IgE, IgG and IgA from the marsupial *Monodelphis domestica*

Maria Aveskogh and Lars Hellman

Department of Medical Biochemistry and Microbiology, University of Uppsala, Uppsala, Sweden

In birds, reptiles and amphibians the IgY isotype exhibits the functional characteristics of both of IgG and IgE. Hence, the gene for IgY most likely duplicated some time during early mammalian evolution and formed the ancestor of present day IgG and IgE. To address the question of when IgY duplicated and formed two functionally distinct isotypes, and to study when IgG and IgA lost their second constant domains, we have examined the Ig expression in a non-placental mammal, the marsupial *Monodelphis domestica* (grey short-tailed opossum). Screening of an opossum spleen cDNA library revealed the presence of all three isotypes in marsupials. cDNA clones encoding the entire constant regions of opossum IgE (ε chain), IgG (γ chain) and IgA (α chain) were isolated, and their nucleotide sequences were determined. A comparative analysis of the amino acid sequences for IgY, IgA, IgE and IgG from various animal species showed that opossum IgE, IgG and IgA on the phylogenetic tree form branches clearly separated from their eutherian counterparts. However, they still conform to the general structure found in eutherian IgE, IgG and IgA. Our findings indicate that all the major evolutionary changes in the Ig isotype repertoire, and in basic Ig structure that have occurred since the evolutionary separation of mammals from the early reptile lineages, occurred prior to the evolutionary separation of marsupials and placental mammals.

Received	24/2/98
Revised	20/5/98
Accepted	26/5/98

Key words: Marsupial / IgG / IgE / IgA / IgY / Evolution

1 Introduction

Neither IgG nor IgE have been identified in reptiles, amphibians or birds and are thus probably unique for mammals [1]. In birds and reptiles antibodies of the IgY isotype exhibit the functional characteristics of both IgG and IgE. Hence, it is probable that IgY duplicated sometime during the early mammalian evolution and formed two separate Ig classes [1]. IgY of birds and reptiles has, like IgE and IgM, four constant domains, while IgG in placental mammals has only three. A sequence comparison between different domains of IgY and IgG shows that it is most likely the second domain of IgY which has been deleted [1]. A short hinge exon is probably the only remaining part of the original exon encoding the second constant domain [1]. Interestingly, IgA in birds also has four constant domains, while the α chain of IgA in placental mammals has only three. Hence, two separate deletion events seem to have occurred during early mammalian evolution. The amphibian *Xenopus laevis* expresses an isotype with similar tissue distribution as mammalian IgA [2]. However, this isotype (IgX) does not seem to be a direct homolog of IgA in mammals and birds. IgX shows, for example, a higher sequence identity with IgM than with mammalian and bird IgA [2]. Interestingly, IgY in the amphibian axolotl (*Ambystoma mexicanum*) has the capacity to be transferred across epithelial surfaces, in that respect resembling mammalian IgA [3]. Consequently, IgY of certain amphibians, which lack IgX, seems to display the effector functions of all three mammalian Ig isotypes (IgG, IgA and IgE).

Marsupials are thought to have separated evolutionarily from the eutherian mammals 140 to 200 million years ago, while the major radiation of the placental mammals probably occurred 140 to 100 million years ago. To further understand the evolution of Ig classes, studies of mammals distantly related to placentals, such as marsupials, may provide useful information. We here present the cloning of the ε chain of IgE, the γ chain of IgG and the α chain of IgA from the marsupial *Monodelphis domestica*, the grey short-tailed opossum.

[I 18062]

0014-2980/98/0909-2738$17.50+.50/0

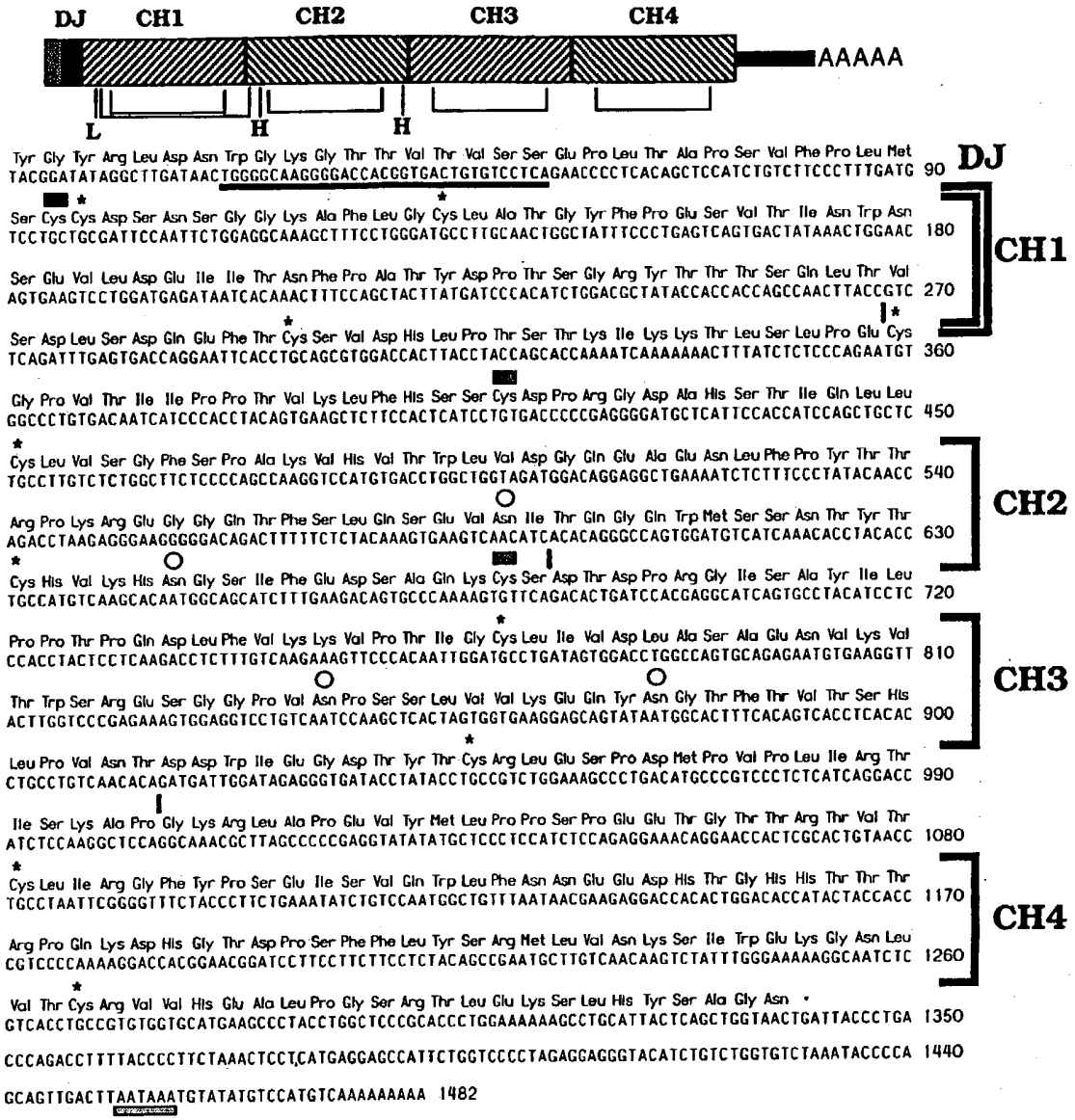
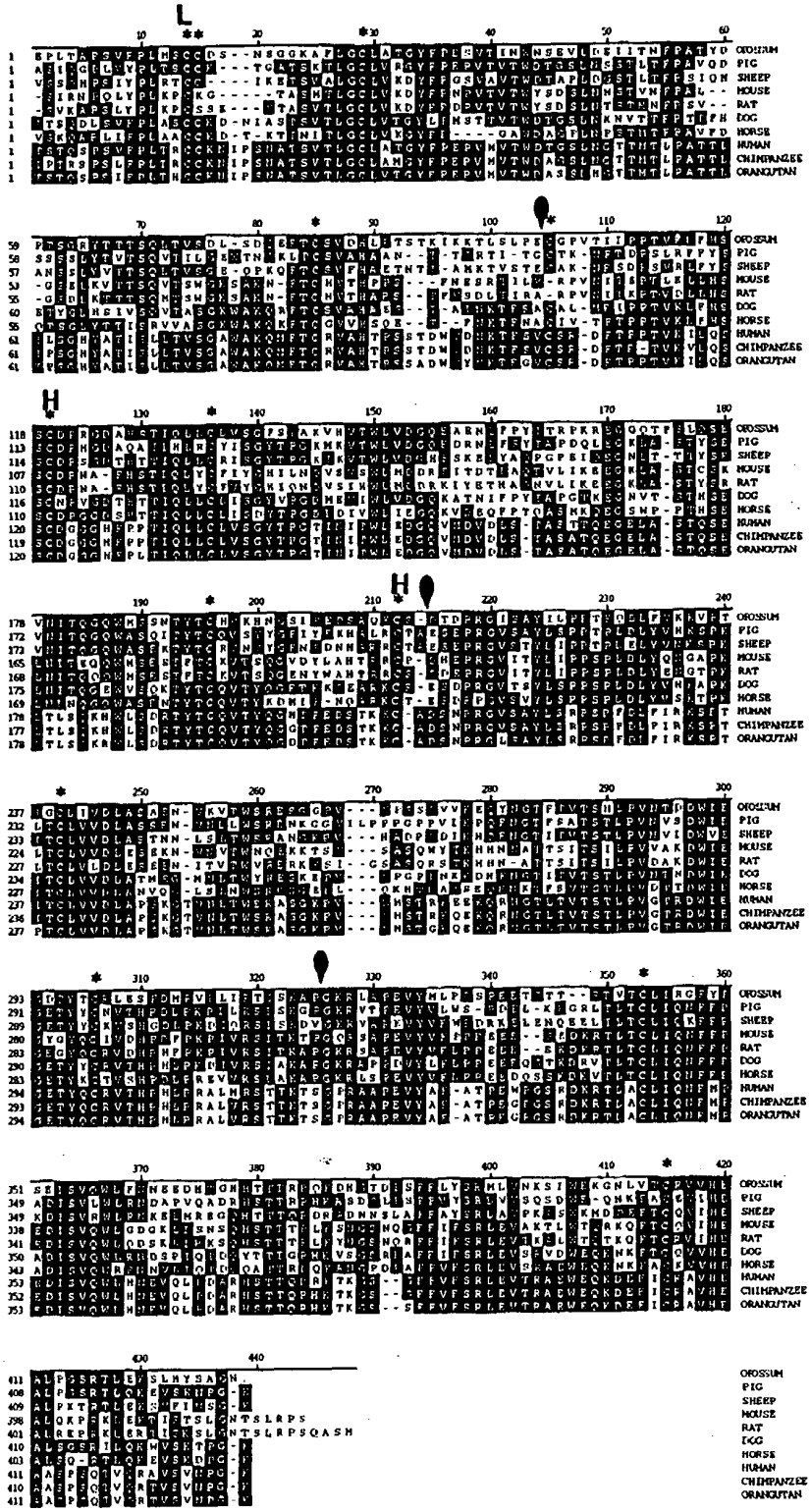


Figure 1. Structure of the opossum ε chain cDNA clone. A schematic map (in scale) of the opossum ε chain cDNA clone is shown at the top of the figure. The striped boxes indicate the four constant regions (CH1–CH4), the joining segment is represented in black and the variable region, together with the D segment, in dark grey. The approximate positions of the intra-chain disulfide bridges have been marked by brackets. An H marks the cysteines involved in putative inter-chain disulfide bridges between the two heavy chain polypeptides, and the L marks the cysteine involved in the putative inter-chain disulfide bridge between the heavy and the light chain polypeptides. The entire sequence of the insert from the original lambda clone, excluding the linker sequences which have been deleted from the 5' and 3' ends, respectively, is depicted below the schematic map. The tentative positions of the exon boundaries, based on the exon-intron structures of mouse, rat, human, and dog ε chain genes, are indicated by small vertical lines. The position of the J segment is highlighted by a thick black line below the corresponding sequence. The cysteines involved in intra-chain disulfide bridges are marked by asterisks, and the different disulfide bridges are marked by thick brackets to the right of the sequence. The amino acid sequence is depicted above the corresponding coding sequence, and potential N-glycosylation sites are indicated by open circles. The nt sequence of opossum IgE has been submitted to the EMBL/GenBank nucleotide sequence databases and has been assigned the accession number AF 035194.


L
H
H
OPOSSUM
PIG
SHEEP
MOUSE
RAT
DOG
HORSE
HUMAN
CHIMPANZEE
ORANGUTAN

◀ *Figure 2.* Amino acid alignment of all presently available ε chain sequences. Amino acid positions that are conserved in at least three of the aligned sequences are represented in black. The cysteine residues involved in inter-chain and intra-chain disulfide bridges are marked by asterisks. The cysteine residues involved in the formation of inter-chain disulfide bridges to the light or heavy chains are indicated by L or H, respectively. The positions of the exon-intron boundaries (in the cases when the gene structure has been determined) are indicated by a pointed black dot above the corresponding sequence.

2 Results

2.1 Cloning and nucleotide sequence analysis of *M. domestica* IgE, IgG and IgA ε, γ and α chains

A PCR-based cloning strategy was adopted for the isolation of cDNA clones encoding the ε and γ chains of opossum IgE and IgG. Initially, a comparison of the nucleotide sequences for the coding regions of the ε and γ chains of a panel of different eutherian IgG and IgE was made. Based on the comparison, two regions with a high degree of sequence conservation for each of these two isotypes were selected for synthesis of degenerate oligonucleotide primers. PCR amplifications, using the degenerate primers, were performed using purified lambda DNA from an amplified cDNA library derived from spleen of the grey short-tailed opossum as template. This resulted in isolation of partial cDNA clones for opossum IgE and IgG. These partial cDNA clones were used as probe to screen the same cDNA library that had been used during the initial PCR amplification. Screening of about 150 000 plaques from this cDNA library with the ε chain probe resulted in approximately 40 positive signals, and screening of approximately 50 000 plaques with the γ chain probe resulted in more than 100 positive signals. Four of the ε chain clones, and eight of the γ chain clones were selected for further analysis. Nucleotide sequence analysis from both ends of the inserts of each clone showed that one of the clones contained the entire coding region for the constant region of the opossum IgE ε chain (Fig. 1) and one clone the entire coding region of a IgG γ chain, including the signal sequence, a variable region, the DJ segments and the entire constant region (Fig. 3).

The opossum cDNA library was then screened with the variable region from the opossum IgG (γ chain) as probe. Following screening of the V region-positive clones with probes for both IgG (γ chain) and IgE (ε chain), 12 IgG- and IgE-negative clones were identified. Using a primer directed against the region spanning the J segment, the nucleotide sequences of the N-terminal region of the constant domains of these 12 clones were determined. From this screening several clones for opossum IgA were identified. Nucleotide sequence analysis from both ends of the inserts of each clone showed that one of the clones contained the entire coding region of a IgA α chain, including the signal sequence, a variable region, the DJ segments and the entire constant region (Fig. 5). The complete nucleotide sequence was determined for the IgE, IgG and IgA ε, γ and α chain clones (Figs. 1, 3 and 5). The corresponding amino acid sequence of the constant region for opossum IgE, IgG and IgA ε, γ and α chains was compared with the previously published amino acid sequences of mammalian IgE, IgG and IgA sequences (Figs. 2, 4 and 6). The result shows that the overall structure of the opossum IgE conforms to the structure of the previously characterized eutherian IgE (Fig. 2). The various cysteine bridges have been conserved, including the second disulfide bridge in the CH1 domain which has been lost in rodent IgE (Figs. 1 and 2). The structural analysis of opossum IgG and IgA shows that the overall structure of these isotypes also conforms to the structure of the previously characterized eutherian IgG, including the presence of three constant domains (Figs. 3–6).

Nucleotide sequence analysis of five independent ε chain clones and eight independent γ chain clones, giving identical sequences within the regions analyzed, indicated that only one IgE and one IgG isotype is expressed in the opossum (data not shown). Nucleotide sequence analysis of two independent α chain clones showed almost identical sequences. A few nucleotide differences between the clones originate most likely from allotypic variations and not from separate isotypes. However, based on only two IgA sequences, we can at present not draw any conclusions on whether one or several IgA isotypes are expressed in the opossum (data not shown).

A comparison was also made between opossum IgG, IgE and IgA and all presently available amino acid sequences for the constant regions of mammalian IgE, IgG and IgA as well as a few IgA and IgY sequences from birds and an amphibian (*A. mexicanum*). The resulting phylogenetic tree shows that the opossum IgE, IgG and IgA form separate branches on the phylogenetic tree, outside of all eutherian IgE, IgG and IgA so far isolated (Fig. 7).

2.2 Heavy chain V region diversity in the opossum

Recent data indicate that only two V gene families (VH1 and VH2) exist in the opossum and that both of these gene families belong to the group III of mammalian V

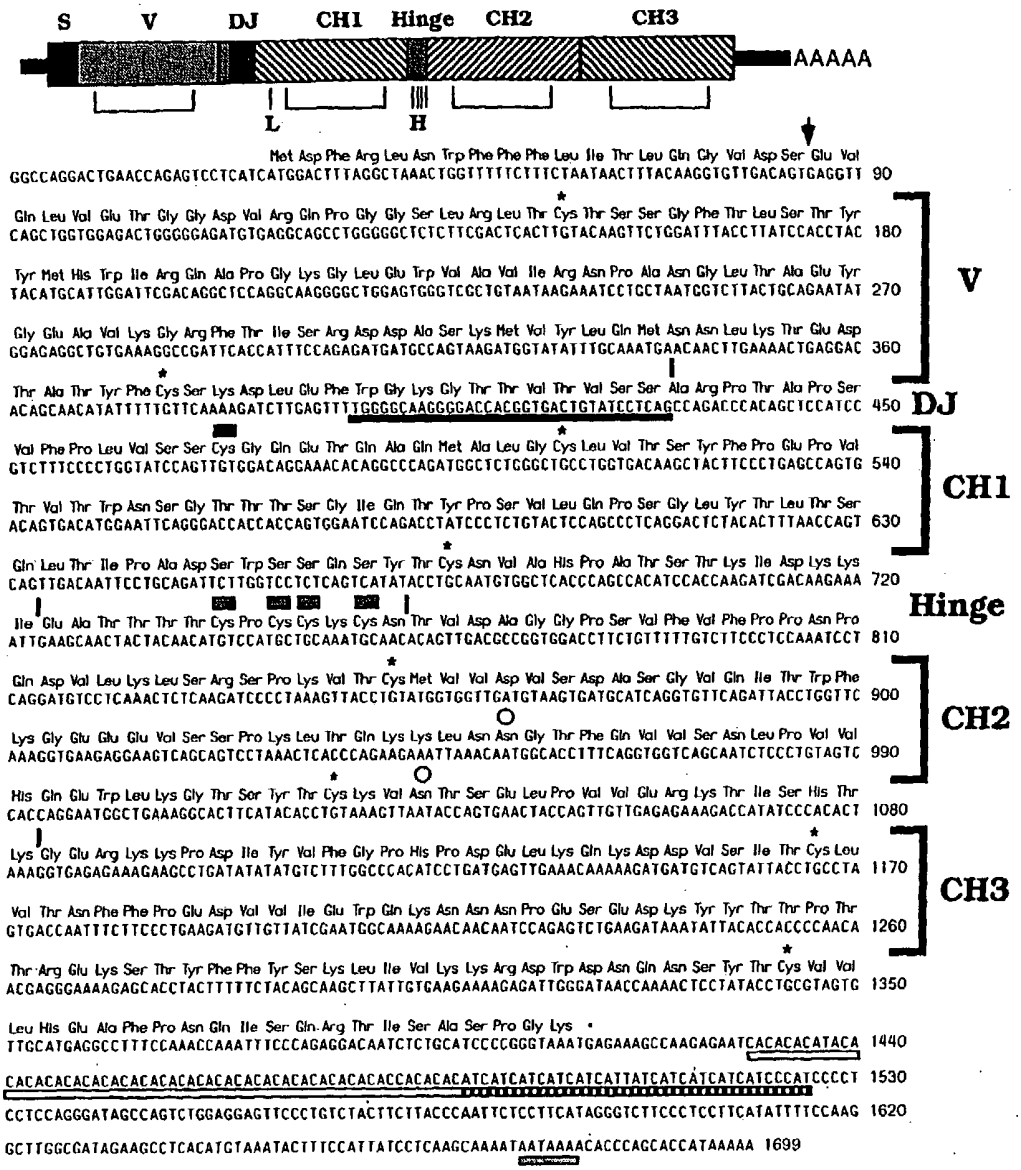


Figure 3. Structure of the opossum γ chain cDNA clone. For details concerning the abbreviations see the legend to Fig. 1. The tentative positions of the exon boundaries, based on the exon-intron structures of mouse γ chain genes, are indicated by small vertical lines. Two regions with repeated sequences were identified in the 3' non-coding region of the mRNA. These two repeats, a CA repeat and a CAT repeat, have been marked by an open and a striped box, respectively. The nucleotide sequence of opossum IgG has been submitted to the EMBL/GenBank nucleotide sequence databases and has been assigned the accession number AF 035195.

gene families [4]. It is not yet known if this apparent limitation in V gene germ-line diversity is compensated for by a particularly large variation in complementarity-determining regions (CDR) of these germ-line genes, or by somatic mutations or by N region diversity in the CDR3.

The variable regions of the opossum γ chain and α chain clones were compared with a panel of these recently published heavy chain V region sequences [4]. The γ chain V region was found to be a new member of the opossum VH2 gene family [4]. Twelve nucleotide differ-

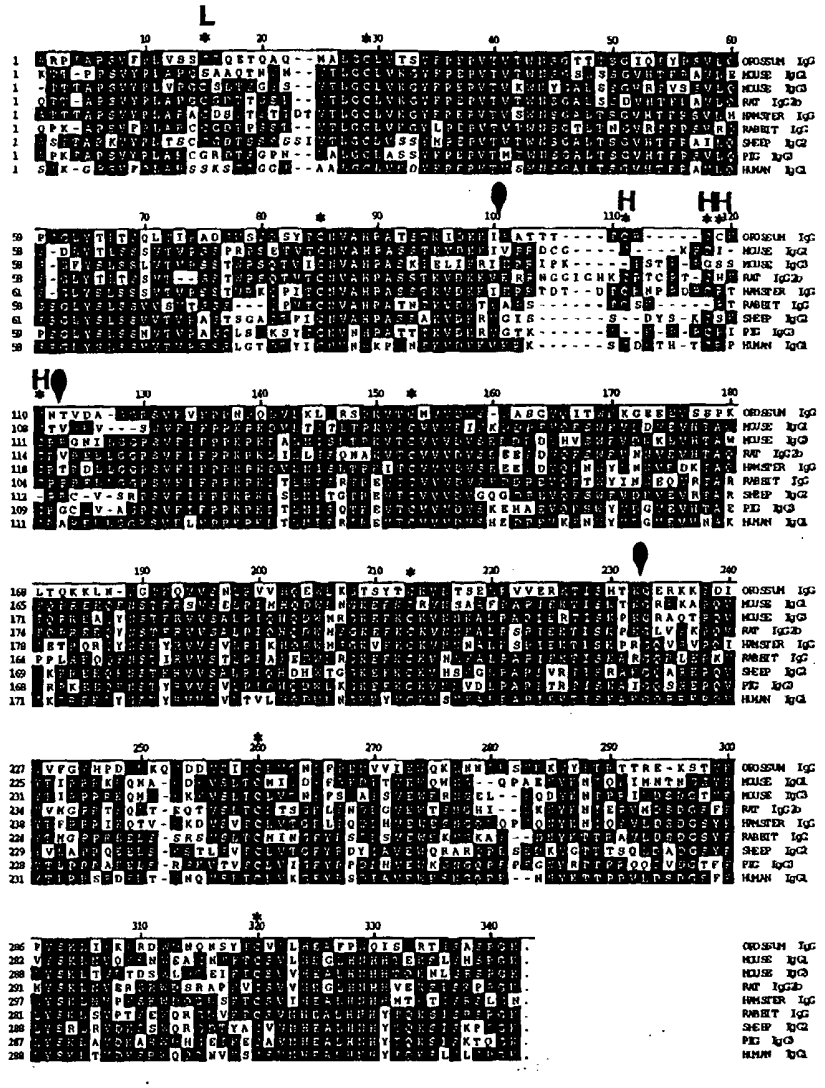


Figure 4. Amino acid alignment of a selected panel of γ chain sequences. For details concerning the abbreviations see the legend to Fig. 2.

ences were found within a 353-nucleotide region, when comparing our sequence with a germ-line VH2 sequence [4]. The variable region of the opossum α chain clone was found to be a new member of the opossum VH1 gene family [4]. Based on these two complete V regions and preliminary data from several additional V region clones (work in progress) we see indications of a relatively large degree of variability in the CDR3 region which may indicate that the variation in the CDR3 may well compensate for the low variability in the germ-line V regions.

3 Discussion

Previous investigations of mammalian Ig evolution have mainly focused on placental mammals. In the present report, we demonstrate for the first time the presence of IgG, IgE and IgA in a mammal distantly related to the placental mammals, the marsupial *M. domestica*. A comparison of the presently available mammalian ε, γ and α chain sequences indicates that only minimal changes in the overall structure have occurred in the last 140–200 million years since the divergence of marsupials and placental mammals [5]. Except for several small

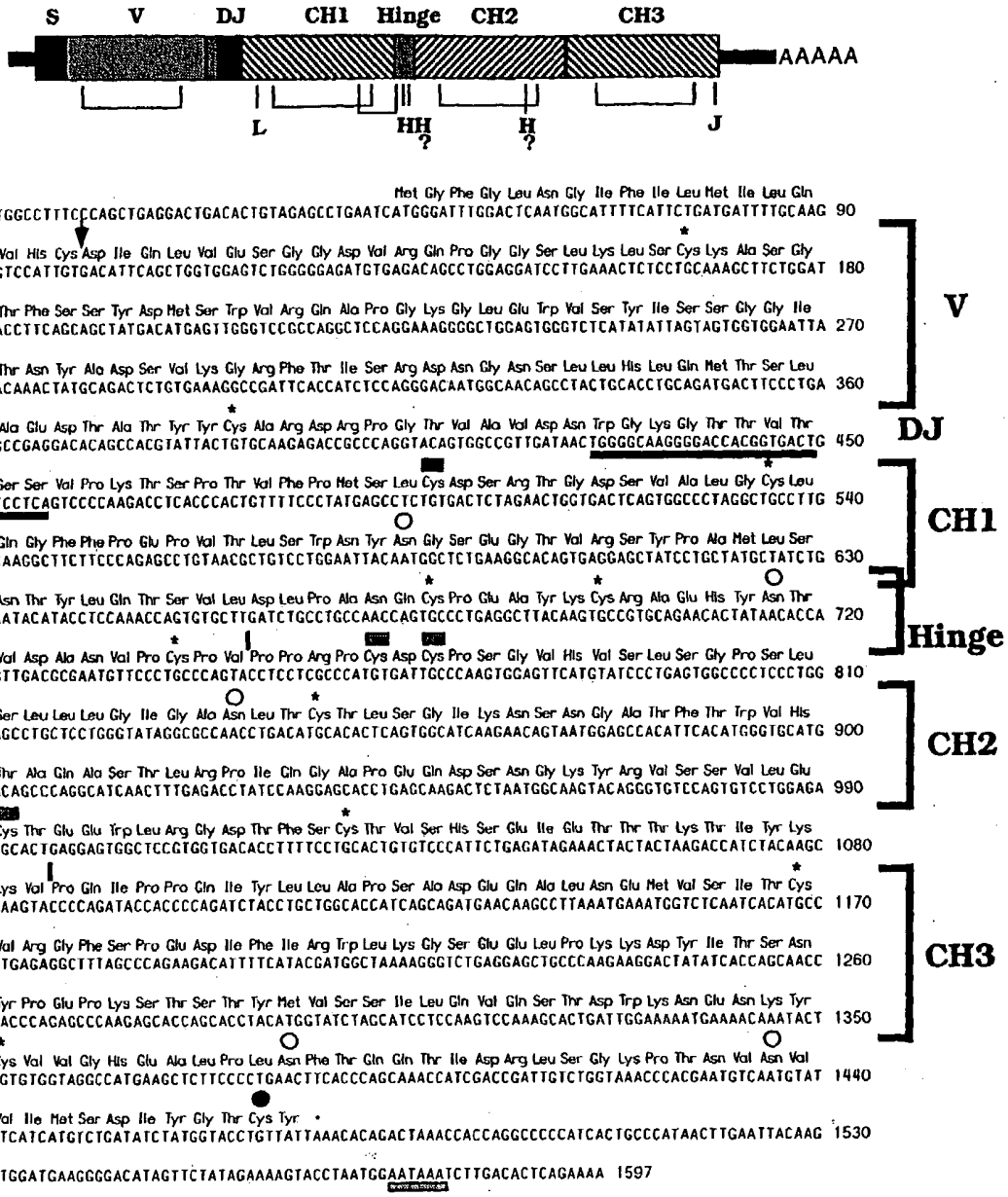


Figure 5. Structure of the opossum α chain cDNA clone. For details concerning the abbreviations see the legend to Fig. 1.

insertions and deletions (one to four full codons in length), only very few changes other than single point mutations have occurred in IgG, IgE and IgA. However, the number of point mutations have been relatively high, resulting in substantial divergence among the various primary nucleotide and amino acid sequences. Both the γ chain of opossum IgG and the α chain of opossum IgA were found to contain only three constant domains (Figs. 3 and 5). This could either be explained by a deletion prior to the evolutionary separation of marsupials and placental mammals or by two independent deletions, one in each of these two major mammalian lineages. The very high overall structural similarity of marsupial and placental IgG and IgA does, however, strongly favor the first alternative. Thus, the most likely explanation is that the second domain of IgG and IgA was

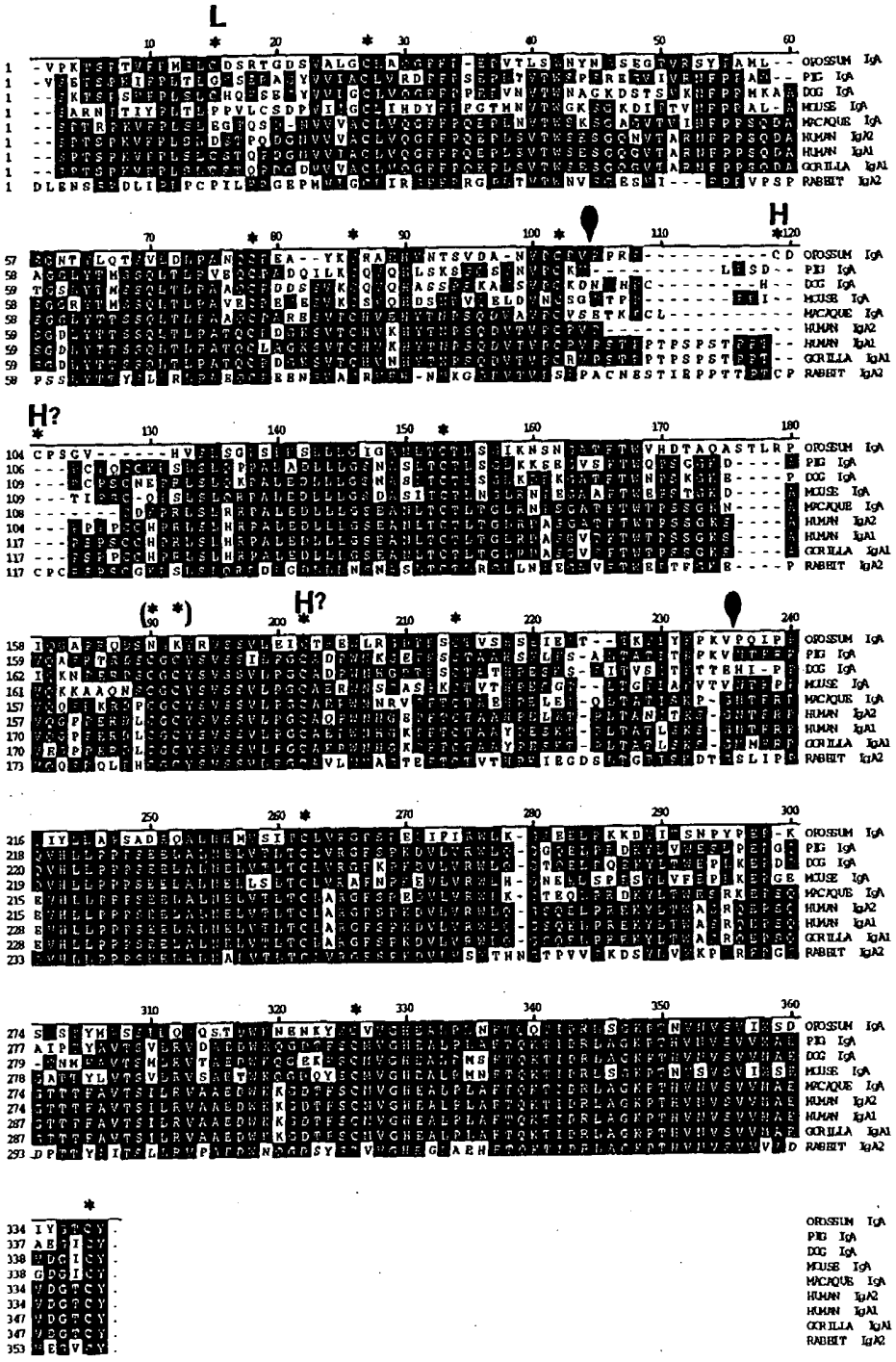


Figure 6. Amino acid alignment of all presently available α chain sequences. For details concerning the abbreviations see the legend to Fig. 2. Two cysteines which are found in all α chain sequences from placental mammals but not in the α chain opossum sequence are marked with asterisks within brackets. The cysteine residues involved in the formation of inter-chain disulfide bridges to the light or heavy chains are indicated by L or H, respectively. Two of these are also marked with a question mark indicating the uncertainty of these cysteine residues. Although less likely, they may instead be involved in an additional intra-chain disulfide bridge (with each other).

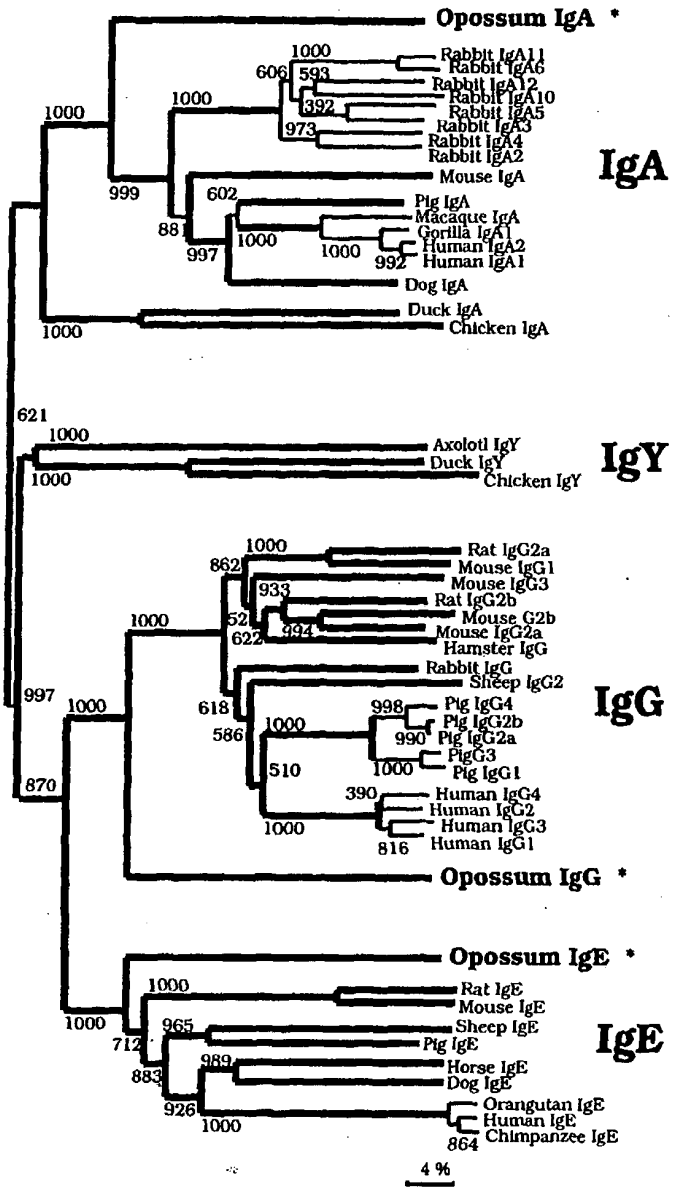


Figure 7. A comparative analysis of the amino acid sequence of the four constant domains from the presently available mammalian IgE, IgG and IgA sequences, two IgA and two IgY sequences from birds and the IgY sequence from the axolotl, presented in a phylogenetic tree. The bootstrap tree is based on 1000 independent comparisons and the bootstrap values are added at each branch point. The sequence distance is represented by the length of the branches according to the scale at the bottom of the figure, which corresponds to a sequence divergence of 4 %.

deleted prior to the evolutionary separation of marsupials and placental mammals. The opossum IgA α chain differs in one major aspect from its eutherian counterparts: it lacks two cysteine residues in the CH2 domain and opossum IgA may thereby lack the second intra-chain disulfide bridge in the CH2 domain found in eutherian IgA (Fig. 6) [6–8]. The significance of this difference for thermal stability or other characteristics of the protein is, however, not yet known. The most striking observation discerned from the comparison of the various ε chain sequences presented in Fig. 2 is that rodent IgE lacks the second intra-chain disulfide bridge in the CH1

domain. This intra-chain disulfide bond is present in all of the other sequences, including the marsupials (positions 15 and 105). Furthermore, both mouse and rat IgE have substantially longer C-terminal ends compared to the other sequences (six and ten amino acids, respectively). These changes have therefore most likely occurred within a rodent ancestor after the major radiation of the placental mammals.

The constant region of opossum IgE was found to contain four potential glycosylation sites (see Figs. 1 and 2). Only one of these sites, the second site in the CH3 domain (position 282 in Fig. 2), is conserved among the various mammalian ε chain sequences, indicating that most of the glycosylation sites in mammalian IgE may be expendable. However, the site corresponding to position 282 in Fig. 2 is conserved in other Ig classes as well, suggesting that glycosylation at this position is of importance for the protein. This position is actually also conserved within the corresponding sequence of the CH2 domain of opossum IgG (Fig. 3).

Many separate gene duplications have presumably occurred during the evolution of the five different Ig classes and subclasses that we see in different mammals of today. By studying the phylogenetic tree presented in Fig. 7 for Ig evolution, we can see that most of these gene duplications leading to expansion of the number of isotypes, within a certain Ig class, have occurred relatively recently and as independent gene duplications within each separate mammalian lineage. The fact that many separate individual gene duplications have resulted in very similar end results is, however, striking. For example rodents as well as man and pig have four to five functional IgG genes. There are, however, exceptions to this pattern. The rabbit and most likely also the opossum has only one IgG gene and the rabbit has at least 13 different genes for IgA.

In contrast to IgG and IgA, only one IgE isotype has been found in all mammals studied, including the marsupials, indicating that there is a selectable advantage of having only one gene for IgE and to have this gene under very stringent control. Because of the potentially harmful effects of IgE, the concentration of plasma IgE has most likely been kept low by a strong control by T cells. During evolution this has probably been easier to maintain by having only one functional IgE gene. One striking example which supports this conclusion is found among the human Ig genes. A relatively recent gene duplication has led to the increase in IgG genes from two to four and an increase in the number of IgA and IgE genes from one to two. Most likely all four of these genes were contained within one single duplicated genomic fragment. Following this duplication all four IgG genes (IgG1 to 4) and both IgA genes (IgA1 and IgA2) have been kept functionally intact. However, the second gene for IgE has been functionally inactivated by a large deletion including parts of the switch region and a region covering the exons for constant domains one and two [9].

In summary, the identification of both IgG and IgE in the opossum shows that the duplication of IgY did occur more than 140–200 million years ago. The presence of only three constant domains in both opossum IgG and IgA also shows that all the major changes in Ig structure that have occurred since the evolutionary separation of mammals from other early reptile lineages had occurred already before the evolutionary separation of marsupials and placental mammals.

4 Materials and methods

4.1 Design of PCR primers

Based on previously published ε chain sequences from mouse, rat, human, orangutan, chimpanzee, dog, horse, pig and sheep [10–9], two regions of particularly high sequence homology were selected as target regions for synthesis of degenerate PCR primers, one in the N-terminal part of the C3 domain and one in the C-terminal region of the C4 domain. The primer sequences are: oligo 212-C3-IgE: GACGAATTCCCCAMRMYYACCTGYCTGGTKKTGGACCTGG; oligo 214-C4B-IgE: GATGGATCCACCTCYAGGGGRCTR WAGRYGAAGAAG.

Based on previously published γ chain sequences from human IgG1, mouse IgG1, rat IgG2a, sheep IgG2, rabbit IgG, hamster IgG, and pig IgG1 [20–27], two regions of particularly high sequence homology were selected as target regions for synthesis of degenerate PCR primers, one in the C-terminal part of the C1 domain and one in the C-terminal region of the C3 domain. The primer sequences are: oligo 226-C1-IgG: GAGGAATTCACMMRSCCASCARCRCCAARGTGGACAAG; oligo 227-C3-IgG: GGAGGATCCTRTRGT GGTTGTGCARRSCCTCRTGYADCAC.

4.2 PCR amplification

Phage particles from approximately 300 µl of an amplified *M. domestica* total spleen cDNA library in λ-ZAP (Stratagene, La Jolla, CA) (3.7×10^7 PFU/ml) were, after addition of 700 µl SM [100 mM NaCl, 10 mM $MgSO_4$, 50 mM Tris pH 7.5 and 0.1 % (w/v) gelatin], precipitated by addition of 0.1 g PEG 6000 and 60 mg NaCl. After the PEG and NaCl had dissolved, the sample was kept on ice for 1 h. The phages were then pelleted by centrifugation for 10 min at 13 000 rpm in an Eppendorf centrifuge. The pellet was dissolved in 300 µl SM and then extracted once with phenol:chloroform (50:50) in the presence of approximately

6 mM EDTA, 0.06 M NaCl and 0.3 % SDS, followed by one extraction with chloroform. The DNA was precipitated with ethanol in the presence of 0.3 M sodium acetate and dissolved in 100 μl double-distilled water. A 5-μl aliquot of the cDNA preparation was used as template in a standard PCR. The PCR was performed in 45 cycles consisting of 40 s at 94 °C, 60 s at 60 °C and 90 s at 72 °C in a MJ Research minicycler.

4.3 Isolation of full-length cDNA for opossum IgG and IgA (γ and α chains) and a cDNA containing the entire constant region of opossum IgE ε chain

A cDNA library of approximately 2×10^6 recombinants was obtained from Stratagene (La Jolla, CA). Approximately 200000 plaques from this amplified cDNA library were spread as a monolayer of the *E. coli* XL1-Blue MRF' strain with a titer of approximately 25000 clones/138-mm plate. The plaques were transferred to Hybond N+ filters (Amersham Int., Amersham, GB). The filters were screened with the fragment derived from the cloned PCR products for opossum IgG γ chain and IgE ε chain, and subsequently washed at high stringency (0.1 × SSC, 0.1 % SDS). Positive plaques were purified, and the pBluescript phagemid of each clone was excised from the lambda ZAP-II vector by helper phage co-infection according to the protocol of the supplier of the lambda ZAP-II library (Stratagene). The different cDNA clones and the cloned PCR product were sequenced from both ends by the use of Sp6 and T7 primers. Based on the obtained sequence, new sequence primers were designed and used to complete the sequence of the entire insert. All nucleotide sequences presented in this study were established by sequencing both strands of the DNA insert.

4.4 Amino acid sequence alignment

The amino acid sequences of the constant regions from a panel of IgE, IgG, IgA, and IgY sequences were compared using the CLUSTAL W program [28]. A bootstrap tree based on 1000 independent calculations was constructed with this program (Fig. 7).

The original references for the amino acid sequences are as follows: human IgE ε chain [13, 14], mouse IgE ε chain [10], rat IgE ε chain [11, 12], chimpanzee IgE ε chain [15], Orangutan IgE ε chain [15], dog IgE ε chain [17], horse IgE ε chain [19], pig IgE [18], sheep IgE ε chain [16], chicken IgA α chain [29], duck IgA α chain (unpublished Genbank accession number, APU27222), mouse IgA [30], rabbit IgA1-13 [31], dog IgA [17], pig IgA [32], macaque IgA [33], gorilla IgA [33], human IgA1 and IgA2 [34], axolotl IgY ψ chain [35], chicken IgY ψ chain [36], duck IgY ψ chain [37], human IgG1 [20], human IgG2 [20, 38], human IgG3 [39], human IgG4 [38, 40], mouse IgG1 [21, 22], mouse IgG2a [41, 42], mouse IgG2b [43, 44], mouse IgG3 [45], rat IgG2a [23], rat IgG2b [23], rat IgG2c [46], sheep IgG2 [24], rabbit IgG [25], hamster IgG [26], and pig IgG1 [27].

Acknowledgement: We would like to thank Dr. Gunnar Pejler and Molly Vernersson for critically reading the manuscript. This investigation was supported by a grant from the Swedish Natural Sciences Research Council (B-BU 9400-301).

5 References

1 **Warr, G. W., Magor, K. E. and Higgins, D. A.,** IgY: clues to the origins of modern antibodies. *Immunol. Today* 1995. **16:** 392–398.

2 **Mussmann, R., Du Pasquier, L. and Hsu, E.,** Is Xenopus IgX an analog of IgA? *Eur. J. Immunol.* 1996. **26:** 2823–2830.

3 **Fellah, J. S., Iscaki, S., Vaerman, J. P. and Charlemagne, J.,** Transient developmental expression of IgY and secretory component like protein in the gut of the Axolotl *(Ambystoma mexicanum). Dev. Immunol.* 1992. **2:** 181–190.

4 **Miller, R. D., Grabe, H. and Rosenberg, G. H.,** VH repertoire of a marsupial *(Monodelphis domestica). J. Immunol.* 1998. **160:** 259–265.

5 **Kardong, K. V.,** Vertebrates, Comparative anatomy, function and evolution. In **Kardong, K.** (Ed.) *The Vertebrate Story.* Wm. C. Brown, Debuque 1995, p 115.

6 **Putnam, F. W., Liu, Y. S. and Low, T. L.,** Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal IgA1 protease, digestion, Fab and Fc fragments, and the complete amino acid sequence of the alpha 1 heavy chain. *J. Biol. Chem.* 1979. **254:** 2865–2874.

7 **Torano, A. and Putnam, F. W.,** Complete amino acid sequence of the alpha 2 heavy chain of a human IgA2 immunoglobulin of the A2m (2) allotype. *Proc. Natl. Acad. Sci. USA* 1978. **75:** 966–969.

8 **Knight, K. L., Martens, C. L., Stoklosa, C. M. and Schneiderman, R. D.,** Genes encoding alpha-heavy chains of rabbit IgA: characterization of cDNA encoding IgA-g subclass alpha-chains. *Nucleic Acids Res.* 1984. **12:** 1657–1670.

9 **Hisajima, H., Nishida, Y., Nakai, S., Takahashi, N., Ueda, S. and Honjo, T.,** Structure of the human immunoglobulin Cε2 gene, a truncated pseudogene: Implications for its evolutionary origin. *Proc. Natl. Acad. Sci. USA* 1983. **80:** 2995–2999.

10 **Ishida, N., Ueda, S., Hayashida, H., Miyata, T. and Honjo, T.,** The nucleotide sequence of the mouse immunoglobulin epsilon gene: comparison with human epsilon gene sequence. *EMBO J.* 1982. **1:** 1117–1123.

11 **Hellman, L., Pettersson, L., Engström, Å., Karlsson, T. and Bennich, H.,** Structure and evolution of the heavy chain from rat immunoglobulin E. *Nucleic Acids Res.* 1982. **10:** 6041–6049.

12 **Steén, M.-L., Hellman, L. and Pettersson, U.,** Rat immunoglobulin E heavy chain locus. *J. Mol. Biol.* 1984. **177:** 19–32.

13 **Flanagan, J. G. and Rabbitts, T. H.,** The sequence of a human immunoglobulin epsilon heavy chain constant region gene, and evidence for three non-allelic genes. *EMBO J.* 1982. **1:** 655–660.

14 **Max, E. E., Battey, J., Ney, R., Kirsch, I. R. and Leder, P.,** Duplication and deletion in the human immunoglobulin ε genes. *Cell* 1982. **29:** 691–699.

15 **Sakoyama, Y., Hong, K. J., Byun, S. M., Hisajima, H., Ueda, S., Yaoita, Y., Hayashida, H., Miyata, T. and Honjo, T.,** Nucleotide sequences of immunoglobulin epsilon genes of chimpanzee and orangutan: DNA molecular clock and hominoid evolution. *Proc. Natl. Acad. Sci. USA* 1987. **84:** 1080–1084.

16 **Engwerda, C. R., Sandeman, R. A., Stuart, S. J. and Sandeman, R. M.,** Isolation and sequence of sheep immunoglobulin E heavy-chain complementary DNA. *Vet. Immunol. Immunopathol.* 1992. **34:** 115–126.

17 **Patel, M., Selinger, D., Mark, G. E., Hickey, G. J. and Hollis, G. F.,** Sequence of the dog immunoglobulin alpha and epsilon constant region genes. *Immunogenetics* 1995. **41:** 282–286.

18 **Vernersson, M., Pejler, G., Kristrersson, T., Alving, K. and Hellman, L.,** Cloning, structural analysis, and expression of the pig IgE ε chain. *Immunogenetics* 1997. **46:** 461–468.

19 **Navarro, P., Barbis, D. P., Antczak, D. and Butler, J. E.,** The complete cDNA and deduced amino acid sequence of equine IgE. *Mol. Immunol.* 1995. **32:** 1–8.

20 **Takahashi, N., Ueda, S., Obata, M., Nikaido, T., Nakai, S. and Honjo, T.,** Structure of human immunoglobulin gamma genes: implications for evolution of a gene family. *Cell* 1982. **29:** 671–679.

21 **Honjo, T., Obata, M., Yamawaki-Katoaka, Y., Kataoka, T., Kawakami, T., Takahashi, N. and Mano, Y.,** Cloning and complete nucleotide sequence of mouse immunoglobulin gamma 1 chain gene. *Cell* 1979. **18:** 559–568.

22 **Sakano, H., Rogers, J. H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. and Tonegawa, S.,** Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA segments. *Nature* 1979. **277:** 627–633.

23 **Bruggemann, M.,** Evolution of the rat immunoglobulin gamma heavy-chain gene family. *Gene* 1988. **74:** 473–482.

24 **Clarkson, C. A., Beale, D., Coadwell, J. W. and Symons, D. B.,** Sequence of ovine Ig gamma 2 constant region heay chain cDNA and molecular modelling of ruminant IgG isotypes. *Mol. Immunol.* 1993. **30:** 1195–1204.

25 **Martens, C. L., Currier, S. J. and Knight, K. L.,** Molecular genetic analysis of genes encoding the heavy chains of rabbit IgG. *J. Immunol.* 1984. **133:** 1022–1027.

26 **Collins, M. and Whitters, M. J.,** Hamster cDNA homologs to the mouse immunoglobulin IgG2 constant and Igh- VH10 genes. *Immunogenetics* 1995. **42:** 225–226.

27 **Kacskovics, I., Sun, J. and Butler, J. E.,** Five putative subclasses of swine IgG identified from the cDNA sequences of a single animal. *J. Immunol.* 1994. **153:** 3565–3573.

28 **Thompson, J. D., Higgins, D. G. and Gibson, T. J.,** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choise. *Nucleic Acids Res.* 1994. **22:** 4673–4680.

29 **Mansikka, A.,** Chicken IgA H chains. Implications concerning the evolution of H chain genes. *J. Immunol.* 1992. **149:** 855–861.

30 **Tucker, P. W., Slightom, J. L. and Blattner, F. R.,** Mouse IgA heavy chain gene sequence: implications for evolution of immunoglobulin hinge exons. *Proc. Natl. Acad. Sci. USA* 1981. **78:** 7684–7688.

31 **Burnett, R. C., Hanly, W. C., Zhai, S. K. and Knight, K. L.,** The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 C alpha genes. *EMBO J.* 1989. **8:** 4041–4047.

32 **Brown, W. R. and Butler, J. E.,** Characterization of a C alpha gene of swine. *Mol. Immunol.* 1994. **31:** 633–642.

33 **Kawamura, S., Tanabe, H., Watanabe, Y., Kurosaki, K., Saitou, N. and Ueda, S.,** Evolutionary rate of immunoglobulin alpha noncoding region is greater in hominoids than in Old World Monkeys. *Mol. Biol. Evol.* 1991. **8:** 743–752.

34 **Flanagan, J. G. and Rabbitts, T. H.,** Arrangement of human immunoglobulin heavy chain constant region genes implies evolutionary duplication of a segment containing gamma, epsilon and alpha genes. *Nature* 1982. **300:** 709–713.

35 **Fellah, J. S., Kerfourn, F., Wiles, M. V., Schwager, J. and Charlemagne, J.,** Phylogeny of immunoglobulin heavy chain isotypes: structure of the constant region of *Ambystoma mexicanum* epsilon chain deduced from cDNA sequence. *Immunogenetics* 1993. **38:** 311–317.

36 **Parvari, R., Avivi, A., Lentner, F. Ziv, E., Tel, O. S., Burstein, Y. and Schechter, I.,** Chicken immunoglobulin gamma-heavy chains: limited VH gene repertoire, com-

binatorial diversification by D gene segments and evolution of the heavy chain locus. *EMBO J.* 1988. **7:** 739–744.

37 **Magor, K. E., Higgins, D. A., Middleton, D. L. and Warr, G. W.,** One gene encodes the heavy chains for three different forms of IgY in the duck. *J. Immunol.* 1994. **153:** 5549–5555.

38 **Ellison, J. and Hood, L.,** Linkage and sequence homology of two human immunoglobulin gamma heavy chain constant region genes. *Proc. Natl. Acad. Sci. USA* 1982. **79:** 1984–1988.

39 **Huck, S., Fort, P., Crawford, D. H., Lefranc, M. P. and Lefranc, G.,** Sequence of a human immunoglobulin gamma 3 heavy chain constant region gene: comparison with the other human C gamma genes. *Nucleic Acids Res.* 1986. **14:** 1779–1789.

40 **Ellison, J., Buxbaum, J. and Hood, L.,** Nucleotide sequence of a human immunoglobulin C gamma 4 gene. *DNA* 1981. **1:** 11–18.

41 **Yamawaki-Kataoka, Y., Miyata, T. and Honjo, T.,** The complete nucleotide sequence of mouse immunoglobin gamma 2a gene and evolution of heavy chain genes: further evidence for intervening sequence-mediated domain transfer. *Nucleic Acids Res.* 1981. **9:** 1365–1381.

42 **Ollo, R., Auffray, C., Morchamps, C. and Rougeon, F.,** Comparison of mouse immunoglobulin gamma 2a and gamma 2b chain genes suggests that exons can be exchanged between genes in a multigenic family. *Proc. Natl. Acad. Sci. USA* 1981. **78:** 2442–2446.

43 **Yamawaki-Kataoka, Y., Kataoka, T. , Takahashi, N., Obata, M. and Honjo, T.,** Complete nucleotide sequence of immunoglobulin gamma 2b chain gene cloned from newborn mouse DNA. *Nature* 1980. **283:** 786–789.

44 **Ollo, R. and Rougeon, F.,** Mouse immunoglobulin allotypes: post-duplication divergence of gamma 2a and gamma 2b chain genes. *Nature* 1982. **296:** 761–763.

45 **Wels, J. A., Word, C. J., Rimm, D., Der-Balan, G. P., Martinez, H. M., Tucker, T. P. and Blattner, F. R.,** Structural analysis of the murine IgG3 constant region gene. *EMBO J.* 1984. **3:** 2041–2046.

46 **Bruggemann, M., Delmastro-Galfre, P., Waldmann, H. and Calabi, F.,** Sequence of a rat immunoglobulin gamma 2c heavy chain constant region cDNA: extensive homology to mouse gamma 3. *Eur. J. Immunol.* 1988. **18:** 317–319.

Correspondence: Lars Hellman, Department of Medical Biochemistry and Microbiology, University of Uppsala, The Biomedical Center, Box 582, S-751 23 Uppsala, Sweden
Fax: +46-18-55 75 79
e-mail: Lars.Hellman@immun.uu.se

Evidence for an early appearance of modern post-switch immunoglobulin isotypes in mammalian evolution (II); cloning of IgE, IgG1 and IgG2 from a monotreme, the duck-billed platypus, *Ornithorhynchus anatinus*

M. Vernersson[1,2], M. Aveskogh[1,2], B. Munday[3] and Lars Hellman[1,2]

[1] Department of Cell and Molecular Biology, University of Uppsala, The Biomedical Center, Uppsala, Sweden
[2] Resistentia Pharmaceuticals AB, Uppsala, Sweden
[3] Department of Human Life Sciences, University of Tasmania, Launceston, Tasmania, Australia

To trace the emergence of the modern post-switch immunoglobulin (Ig) isotypes in vertebrate evolution we have studied Ig expression in mammals distantly related to eutherians. We here present an analysis of the Ig expression in an egg-laying mammal, a monotreme, the duck-billed platypus (*Ornithorhynchus anatinus*). Fragments of platypus IgG and IgE cDNA were obtained by a PCR-based screening using degenerate primers. The fragments obtained were used as probes to isolate full-length cDNA clones of three platypus post-switch isotypes, IgG1, IgG2, and IgE. Comparative amino acid sequence analysis against IgY, IgE and IgG from various animal species revealed that platypus IgE and IgG form branches that are clearly separated from those of their eutherian (placental) counterparts. However, the platypus IgE and IgG still conform to the general structure displayed by the respective Ig isotypes of eutherian and marsupial mammals. According to our findings, all of the major evolutionary changes in the expression array and basic Ig structure that have occurred since the evolutionary separation of mammals from the early reptile lineages, occurred prior to the separation of monotremes from marsupial and placental mammals. Hence, our results indicate that the modern post-switch isotypes appeared very early in the mammalian lineage, possibly already 310–330 million years ago.

Received	4/12/01
Revised	8/5/02
Accepted	28/5/02

Key words: Monotreme / IgG / IgE / IgY / Evolution

1 Introduction

The combinatorial immune system defined by bona fide immunoglobulins and T cell receptors seems to have emerged at the appearance of the jawed vertebrates [1–3]. Due to the absence of its components in invertebrates as well as agnathans the combinatorial immune system has been proposed to have appeared in a relatively short period of evolutionary time, possibly in a single step catalyzed by horizontal transfer of recombination genes from prokaryotes into the vertebrate genome. This event has been described as a "Big Bang" in molecular evolution [2, 3]. Since this "Big Bang", the complexity of the combinatorial immune system has increased gradually throughout vertebrate evolution. Bony and cartilaginous fishes express primarily IgM, whereas eutherian (placental) mammals express up to five Ig classes (IgM, IgD, IgG, IgE and IgA). Fish B cells apparently do not undergo isotype switching [4]. However, depending on the species, an IgD-like Ig, IgNAR, IgNARC, IgW or IgR (IgX) may also be expressed, either through differential splicing (the IgD-like Ig) or their presence on loci separate from that of IgM, reviewed in [4].

Amphibians, reptiles and birds all express IgY, the first confirmed post-switch isotype to appear in evolution [5, 6]. Hence, the mechanism of isotype switching seems to have emerged with the tetrapods, or possibly in the putative ancestors of tetrapods, the lungfish or the coelacanth (*Latimeria*) [6]. Early studies in the lungfish have identified a second highly abundant serum Ig in addition to IgM that shows similarities with the truncated form of IgY found in ducks [7–9]. However, further studies in the lungfish and the coelacanth are required to establish at which evolutionary stage the first post-switch isotype did

[I 22682]

0014-2980/02/0808-2145$17.50+.50/0

appear. In addition, such analyses may help to resolve the origin of tetrapods [10].

IgY shares effector functions characteristic of both mammalian IgG and IgE. Hence, an IgY-like ancestor probably duplicated sometime early in mammalian evolution and formed two separate Ig classes, IgG and IgE [5, 6].

Like IgE and IgM, IgY of birds and amphibians has four constant domains, whereas IgG in eutherian mammals and marsupials has only three. Sequence comparisons among different domains of IgY and IgG suggest that the second domain of the ancestral IgY was deleted in what evolved into the IgG of mammals [5]. A short hinge exon found in mammalian IgG seems to be the only remaining part of the original exon encoding the second constant domain in the ancestral IgY [5].

Frogs (*Xenopus laevis*) but not axolotls, express one additional post-switch isotype called IgX (different from IgX of fish), that probably arose later than IgY [6]. *Xenopus* IgX appears to have similar distribution and effector functions as IgA in mammals and birds, however, it is probably not the direct homologue of mammalian IgA [11].

Most earlier studies of immunoglobulin expression in mammals have focused on eutherian mammals, where all investigated species express IgM, IgG, IgE and IgA. Some eutherian mammals, such as humans and mice, express IgD as well. Apart from the eutherians, there are two other extant mammalian lineages, the marsupials (metatheria) and the monotremes (prototheria).

Marsupials are represented by more than 270 living species, whereas the monotremes, a group of egg-laying mammals, are represented by only three, the duck-billed platypus and two echidna species. The monotremes have been regarded as "reptile-like" or "primitive" mammals simply because they lay eggs. However, they do possess several of the major mammalian features including a well-developed fur coat, a single bone in the lower jaw, three bones in the middle ear and mammary glands. Histological studies show that the spleen, thymus and gut-associated lymphoid tissues in the platypus are well developed and comparable in histological structure to those in therian mammals (eutherians and marsupials) [12]. However, in sites where lymph nodes would be expected in therian mammals, monotremes were found to have lymphoid nodules that resemble the jugular bodies of the amphibians, which indicates that monotremes have a somewhat more primitive immune system [12].

The evolutionary separation of the marsupials from the placental mammals is generally dated back to approximately 130 to 175 million years ago (Belov et al., "echidna IgM", submitted for publication) [13, 14], whereas the major radiation of the placental mammals probably occurred 60 to 120 million years ago [14]. The evolutionary relationship between marsupials and monotremes has not yet been fully resolved. The Marsupionata hypothesis supported mainly by mitochondrial data and DNA-hybridization studies suggests that marsupials and monotremes are sister lineages [15–18]. The opposing Theria hypothesis proposes that the separation of the monotremes from the marsupial and eutherian mammals occurred much earlier in evolution. The Theria hypothesis initially emerged from morphological data and has more recently gained additional support from paleontological data [19, 20] and a number of nuclear gene sequences (Belov et al., "echidna IgM" and "echidna IgA", both submitted for publication) [13, 21–24]. The nucleotide substitution rate in mitochondrial DNA is generally more rapid and the tendency for repeated mutations in the same site is higher than what is observed in nuclear DNA [25]. Consequently, time estimates based on mitochondrial data may be less accurate over large evolutionary distances. In fact, Springer and colleagues [25] showed that mitochondrial genes have less resolving power than certain nuclear exons in resolving deep-level mammalian clades. Studies based on IgM (Belov et al., "echidna IgM", submitted for publication) and protamine sequences estimate the time of divergence of the monotremes from the other mammalian lineages to around 150 to 170 million years ago [13]. This estimate fits quite well with dating based on fossil records and anatomical features [19, 20, 26].

In a previous study of marsupial immune functions, we found that the American short-tailed gray opossum expresses four Ig isotypes, IgM, IgG, IgE and IgA, which all conform to the basic structure of the corresponding isotypes in placental mammals [27]. The findings in the opossum revealed that all of the major changes in the Ig isotype pattern that emerged after the separation of the mammals from the early reptile lineages had occurred prior to the separation of marsupials from placental mammals. Hence, the duplication of IgY and the deletion of the second constant domain in the ancestral IgG and IgA probably took place more than 130 million years ago.

To further trace the evolution of modern post-switch isotypes and the appearance of the full range of functionally separated Ig isotypes found in therian mammals, referred to as "a second Big Bang" by Marchalonis and colleagues [3], we initiated a study of Ig isotypes in monotremes. As a first step, we describe here the cloning of IgE, IgG1 and IgG2 from the duck-billed platypus, *Ornithorhynchus anatinus*.

2 Results

2.1 Cloning and nucleotide sequence analysis of the ε- and γ-chains of platypus IgE, IgG1 and IgG2

A PCR-based cloning strategy was adopted to isolate cDNA clones encoding partial platypus IgE and IgG heavy chains. Nucleotide sequence comparisons of the ε- and γ-chain sequences from a panel of different mammals had been conducted earlier [27]. Based on these comparisons, two regions with particularly high sequence conservation within the respective isotypes were selected as target regions for synthesis of degenerate oligonucleotide primers. The degenerate primers were used in PCR amplifications using double-stranded cDNA from Australian platypus splenocytes as template. This resulted in the isolation of one partial cDNA clone each for platypus IgE and IgG. The clone generated with the ε-chain primers was approximately 470 base pairs (bp), encoding a region stretching from the CH3 to CH4 central regions of the ε-chain. The partial IgG clone was approximately 540 bp stretching from the N-terminal part of the CH2 domain to the C-terminal part of the CH3 domain.

The partial platypus IgE and IgG cDNA clones were used as probes to screen a platypus spleen cDNA library. Screening of about 130, 000 plaques (16,250 plaques/filter) resulted in an average of 16 and 115 positive signals/filter using the ε- and γ-chain probes, respectively. Four positive clones, two from each screening, were selected for further analysis. The complete nucleotide sequences of the ε- and γ-chains were determined from one full-length clone of the respective isotypes (GeneBank AY055780 and AY055781).

Upon further screening of the library using a full-length IgG clone as probe, two clones for a novel γ-chain isotype were isolated and sequenced, together with nine additional clones of the previously isolated γ-chain isotype. The initially isolated γ-chain and the novel γ-chain isotype were designated IgG1 and IgG2, respectively. The complete nucleotide sequence of one full-length γ2-chain was determined (GeneBank AY055782).

The amino acid sequence of the variable domains of the full-length ε, γ1 and γ2 clones are shown in Fig. 1 and the amino acid sequence of the constant domains of IgE and the two IgG isotypes are shown in Fig. 2 and 3, respectively.

Following the isolation of full-length clones for IgG1, IgG2 and IgE, the platypus cDNA library was screened with a variable region (V-region) probe. Based on previous screenings using IgG1, IgG2 and IgE constant-region-specific probes, IgG and IgE negative V-region positive clones were isolated. To avoid abolition of closely related isotypes the filters were washed at high stringency following all constant region hybridizations. Thirteen of the V-region-positive, γ- and ε-chain-negative clones were subcloned and sequenced using a primer directed against the region spanning the J segment. All of these thirteen clones were clearly identified as α-chains of two different IgA-like isotypes: three α1-chains (IgA1) and ten α2-chains (IgA2). Hence, no additional IgG or IgE isotypes were identified and we failed to identify any IgM or IgD like isotypes. IgM is present in all mammals investigated so far, including another monotreme, the Echidna (Belov et al., submitted for publication), and therefore likely to be present also in the platypus. However, a similar screening strategy was used to isolate the various Ig isotypes expressed by the opossum (*Monodelphis domestica*) [27, 28] and in this screening 9 out of 12 IgG- and IgE-negative clones were identified as IgM, whereas the remaining 3 turned out to be IgA (unpublished observations). Therefore, our failure to identify IgM in platypus is an intriguing subject for ongoing investigation.

The complete primary structure of platypus IgA1 and IgA2 will be presented elsewhere (manuscript in preparation).

2.2 Sequence comparisons and structural analysis of the ε- and γ-chains of the platypus IgE and IgG

The complete amino acid sequences of the constant regions of platypus IgE, IgG1 and IgG2 were compared with amino acid sequences of the corresponding isotypes of other mammals (Fig. 2 and 3). The overall structures of the platypus IgE and IgG conform to the structures of the previously characterized eutherian and marsupial IgE and IgG.

The amino acid sequence of platypus IgE reveals that the cysteine residues involved in the various disulfide bridges that are conserved in most mammalian IgE have been conserved also in platypus IgE. Furthermore, both of the putative inter-chain disulfide bridges present in the CH2 domain of IgE in all species analyzed so far are conserved also in platypus IgE (Fig. 1 and 2). Five potential N-linked glycosylation sites were identified in the constant region of platypus IgE: two in the C2 domain, one in the C3 domain and two in the C4 domain (Fig. 2). The site in the CH3 domain is conserved among all ε-chain sequences isolated so far, suggesting that glycosylation

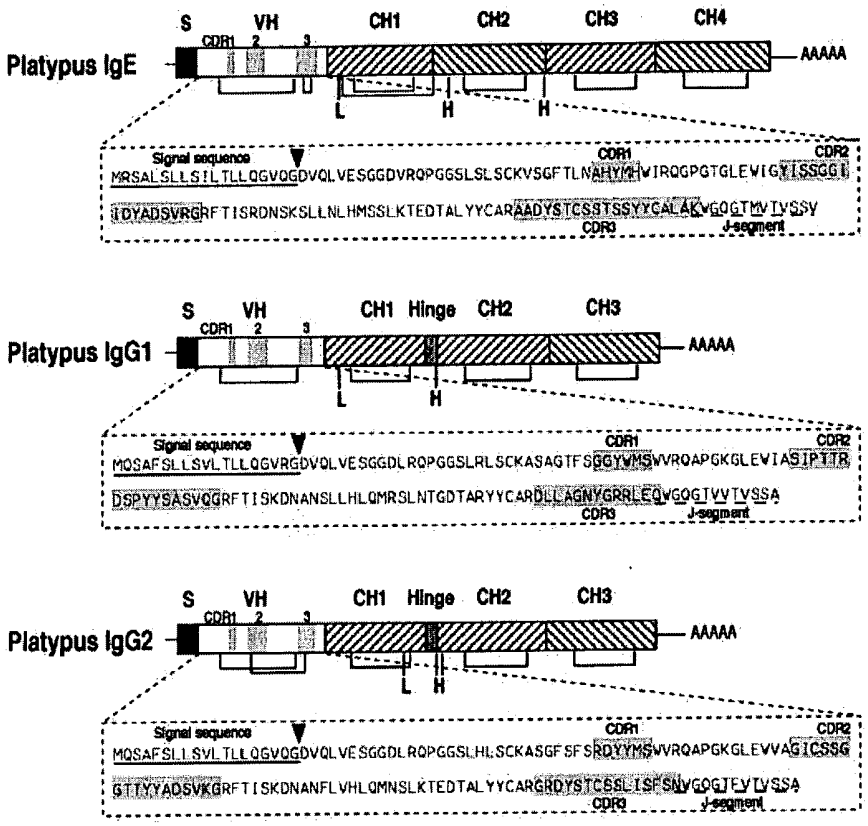


Fig. 1. Schematic maps (in scale) of the platypus ε, γ1 and γ2 heavy-chain clones and the amino acid sequence of the variable domains encoded by the V, D and J gene segments of the respective full-length clones. The signal sequences (S) are represented in black, the variable domain (VH) is represented in white with light gray regions denoting the positions of the complementary-determining regions (CDR), the constant domains are illustrated by striped boxes and the hinge region (in IgG) is indicated in dark gray. Brackets indicate the positions of putative intra-chain disulfide bridges. Cysteine residues involved in putative inter-chain disulfide bridges connecting the two heavy chains and the heavy chain with the light chain are denoted by H and L, respectively. The amino acid sequences of the variable domains are shown below the schematic maps. A small arrow marks the position of the signal sequence cleavage site and the amino acid residues corresponding to CDR1, 2 and 3 are indicated by light gray boxes. CDR borders were set according to Kabat et al. [29].

at this position is of importance for protein structure and/or solubility (position 276 in Fig. 2).

Like the previously characterized eutherian and marsupial IgG, the heavy chains of the platypus IgG isotypes, IgG1 and IgG2, contain three constant domains (Fig. 1 and 3). However, although the platypus IgG conform to the general IgG structure, some important differences were observed. For instance, in platypus IgG2 the cysteine residue forming the disulfide bridge that connects the heavy (H) chain with the light (L) chain is positioned in the C-terminal region of the CH1 domain (Fig. 1 and 3). In contrast, in most, if not in all, previously characterized mammalian IgG molecules, this cysteine residue occurs in the N-terminal part of the CH1 domain (Fig. 3). One potential inter-chain disulfide bridge-forming cysteine residue is found in the hinge region of platypus IgG1, whereas two cysteines are present in the same region of IgG2 (Fig. 1 and 3). On the whole, the two IgG isotypes found in platypus display a remarkable divergence in sequence compared to IgG isotypes of other mammalian species.

Sequence analysis of the V-regions also revealed some interesting features. In addition to the canonical cysteine residues, the V-regions of both the IgG2 and the IgE clone have two extra cysteine residues located in the complementarity-determining region (CDR)2 and/or CDR3 loop structures (Fig. 1). Considering the high reactivity of thiol groups and the structural flexibility of the loop structures, additional intra-chain disulfide bridges probably occur in these V-regions. Moreover, the addi-

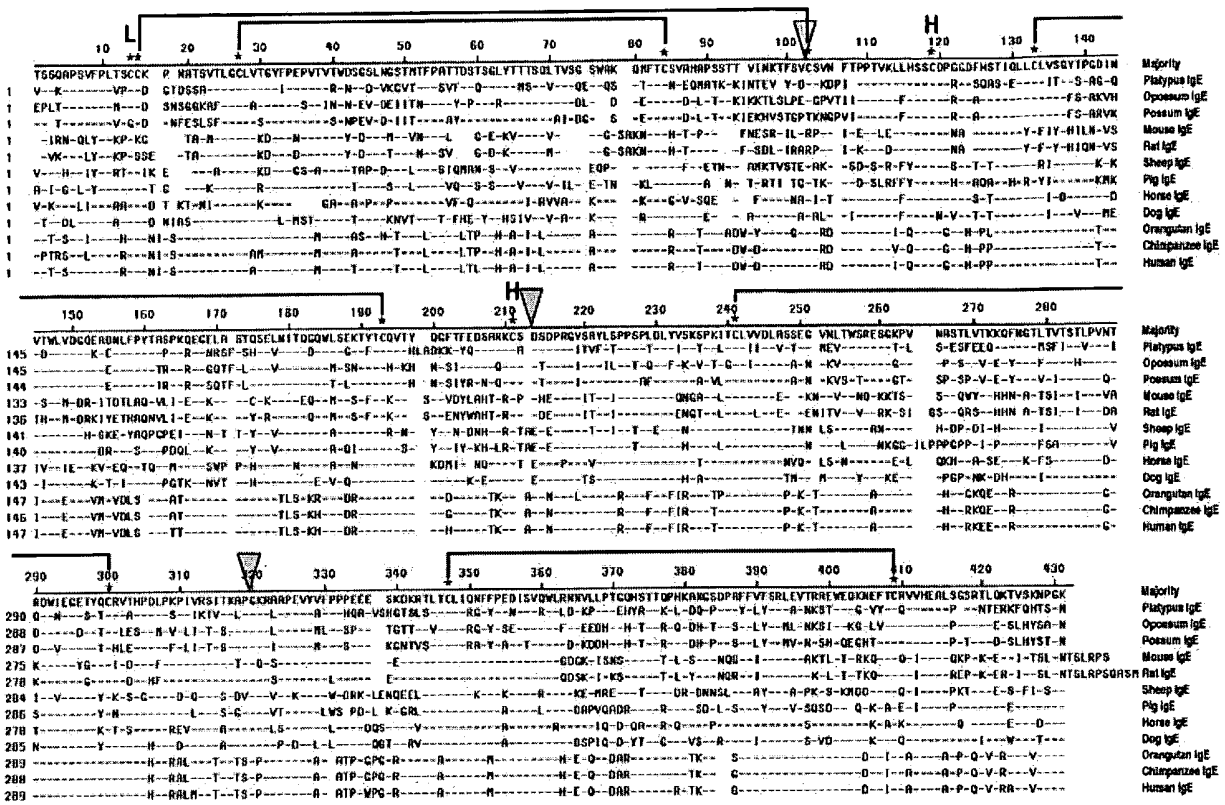


Fig. 2. Amino acid sequence alignment of a selected panel of ε-chains. The top sequence is generated by the amino acid residues that are represented in the majority of the sequences. Residues that match with the consensus sequence are indicated by a dash. The positions of the exon/intron boundaries (based on sequences in which gene structure has been determined) are indicated by arrowheads above the sequence. All cysteine residues involved in putative inter-chain and intra-chain disulfide bridge formations are marked by asterisks and the intra-chain disulfide bridges are indicated by brackets. Cysteine residues involved in the formation of inter-chain disulfide bridges connecting the heavy chains and the heavy chain with the light chain are denoted by H and L, respectively.

tional cysteine residue in the CDR2 of platypus IgG2, residue 52 according to [29], appears to face the CDR3 [30]. A similar arrangement of cysteine residues in the heavy chain V-regions of the llama was suggested to be compatible with a disulfide bridge between CDR2 and CDR3 [31]. The presence of intra-CDR3 disulfide bridges has been indicated in the heavy chain V-regions of shark [32], cow [33] and members of Camelidae [31, 34]. Hence, a disulfide bridge may form between the CDR2 and CDR3 of the IgG2 clone and an intra-CDR3 cysteine bridge probably occurs in the IgE clone (Fig. 1). A more detailed analysis of the V-gene repertoire in platypus revealed that additional cysteine residues that form putative disulfide bridges, either between the CDR2 and CDR3 loop structures or within the CDR3, were found in 9 of 25 clones analyzed [35].

A phylogenetic tree was generated by comparing the complete amino acid sequence of the platypus IgE and IgG constant regions with those of a selection of the presently available mammalian IgE and IgG sequences, along with two IgY sequences from birds (Fig. 4). The platypus IgE and IgG form separate branches, outside of all the eutherian IgE and IgG. Strikingly, the branch for the platypus IgG isotypes occurs outside of both marsupial and eutherian IgG, whereas the platypus IgE sequence indicates that monotremes and marsupials are sister lineages.

2.3 Relative expression levels of platypus IgE, IgG1 and IgG2 in the spleen of a free-living Tasmanian platypus and Northern blot analysis

The relative expression levels of the various isotypes in the spleen of a free-living platypus were estimated by screening the non-amplified spleen cDNA library with

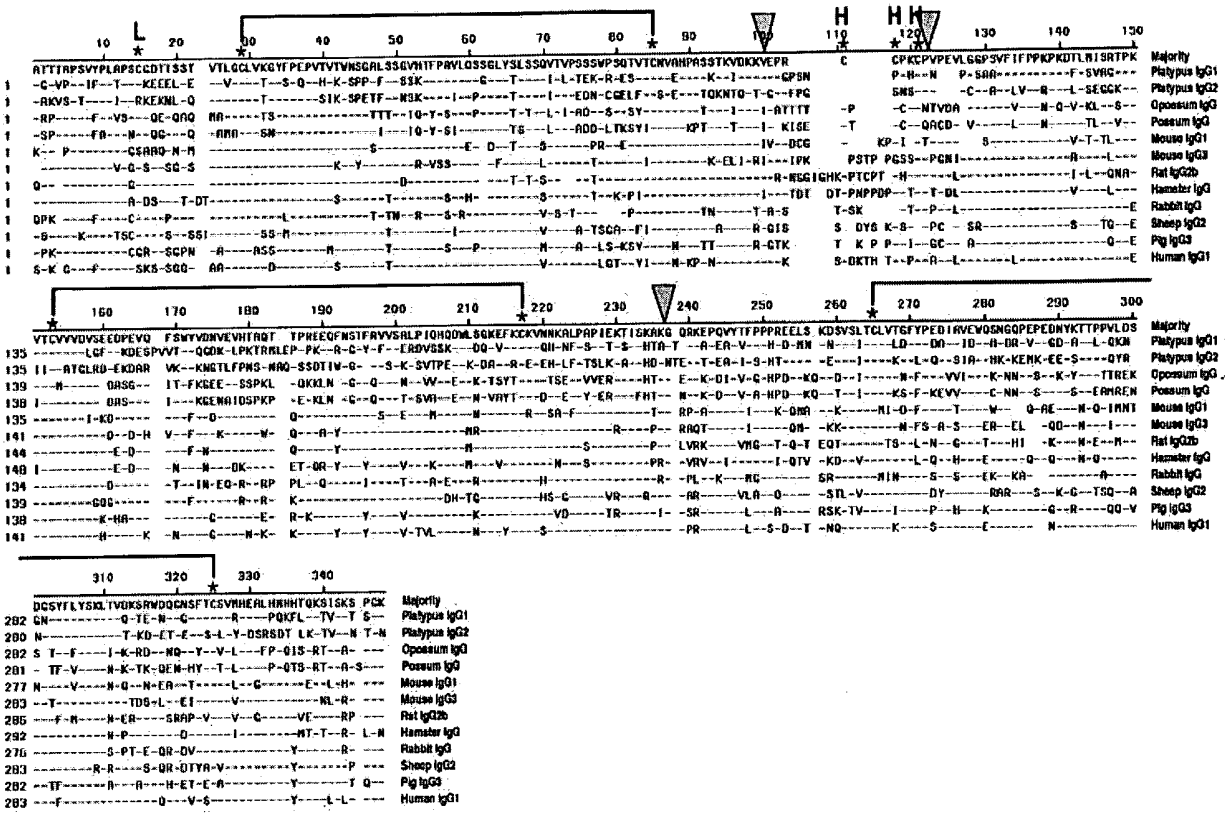


Fig. 3. Amino acid alignment of a selected panel of γ-chain sequences. The top sequence is generated by the amino acid residues that are represented in the majority of the sequences. Residues that match with the consensus sequence are indicated by a dash. The positions of the exon/intron boundaries (based on sequences in which gene structure has been determined) are indicated by arrowheads above the sequence. All cysteine residues involved in putative inter-chain and intra-chain disulfide bridge formations are marked by asterisks and the intra-chain disulfide bridges are indicated by brackets. Cysteine residues involved in the formation of inter-chain disulfide bridges connecting the heavy chains and the heavy chain with the light chain are denoted by H and L, respectively.

isotype specific probes. No cross-hybridization to other isotypes was observed with the isotype-specific probes. The screenings revealed that IgE, IgG1 and IgG2 are expressed at relative levels of 10, 60 and 1. Hence, IgG1 seems to be the predominant plasma IgG isotype in the platypus, whereas IgG2 is much less abundant. IgG1 is expressed at six times the level of IgE and the expression level of IgG2 is only one tenth of that observed for IgE. However, it should be emphasized that these values are based on the relative abundances of mRNA and need not reflect the serum levels of the various Ig isotypes.

None of the 28 analyzed Ig clones contained exons corresponding to the membrane region, indicating that the expression levels of the membrane-bound forms are comparatively low. A Northern blot analysis using isotype-specific probes was performed to estimate the relative abundance of transcripts encoding membrane-bound versus secreted forms of the platypus IgE, IgG1 and IgG2 isotypes, respectively. However, no transcripts encoding the membrane-bound forms of the three isotypes were detected (data not shown). Based on our results we estimate that in the spleen less than 5% of the total mRNA pool for each of the respective isotypes corresponds to the membrane-bound forms, which appears to be in agreement with the situation in other mammals.

3 Discussion

We have demonstrated the presence of IgE and two separate IgG isotypes, IgG1 and IgG2, in a monotreme *Ornithorhynchus anatinus*, the duck-billed platypus.

Sequence comparisons of the platypus isotypes and a selection of the presently available mammalian ε- and γ-

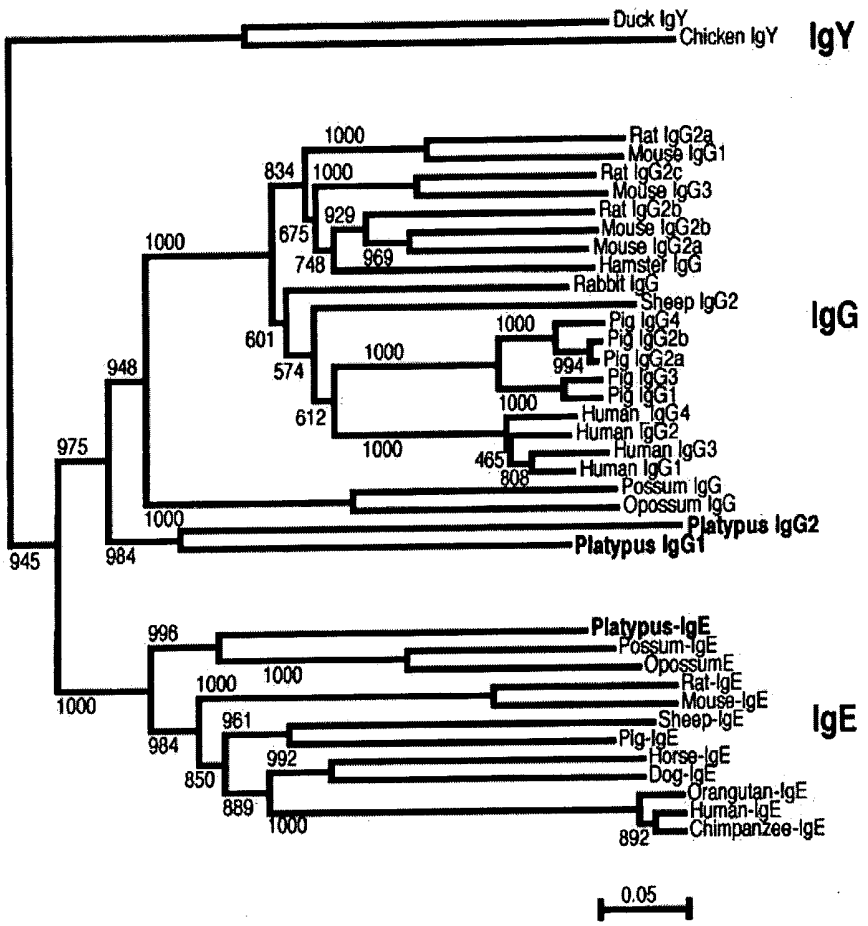


Fig. 4. A comparative analysis of the amino acid sequence of the entire constant region of a selected panel of mammalian IgE and IgG sequences and two IgY sequences from birds presented in the form of a phylogenetic tree. The bootstrap tree is based on 1,000 independent comparisons and the bootstrap values are added at each branch point. The sequence distance is represented by the length of the branches according to the scale at the bottom of the figure, which corresponds to a sequence divergence of 5%.

chain sequences indicate that only minor changes in the overall structures of IgE and IgG have occurred since the evolutionary separation of the monotremes from the other mammalian lineages (Fig. 2 and 3). Apart from several small insertions and deletions (1–4 full codons) and a few changes in the positions and numbers of disulfide bridges and putative glycosylation sites, few changes other than single point mutations have occurred. However, due to the relatively high number of point mutations, the divergence among the various primary nucleotide and amino acid sequences is substantial (Fig. 2 and 3).

The most striking differences among the various ε-chain sequences in Fig. 2 occur in rodent IgE, which has lost one of the intra-chain disulfide bridges in the CH1 domain (position 15 and 103) and has substantially longer C terminus than all other mammalian IgE. These characteristic features probably appeared in a common rodent ancestor after the major radiation of the placental mammals. Interestingly, the recent finding that this intra-chain disulfide bridge is absent also in a marsupial ε-chain, that of the Australian brushtail possum, suggests that it has been lost at least twice during mammalian evolution [36].

The mRNA level for IgE was surprisingly high in the platypus spleen. Only a sixfold difference in mRNA levels was observed between the predominant serum antibody in platypus, IgG1, and IgE. This is intriguing, since, in general, the serum concentration of IgE is significantly lower than that suggested by the mRNA level. In human adults, the average serum level of IgG is 9.5–12.5 mg/ml, whereas IgE is present at concentrations that are 30,000

times lower (10–400 ng/ml). The mRNA levels presented here are based on studies in a free-living platypus. Hence, increased expression of IgE as a result of parasitic infections may account for some elevation in the mRNA level. Furthermore, in humans, the turnovers of IgG and IgE differ quite remarkably. Serum IgG has a half-life of about 3 weeks (23 days), whereas the half-life of IgE in the circulation has been estimated at 2–2.5 days [37]. The expression level of IgE is thus much higher than can be expected on the basis of serum levels alone.

Several findings indicate a common phylogenic history of the avian and amphibian ν-chains and the mammalian ε- and γ-chains [5, 6, 9]. In analogy with IgG in mammals, IgY is the major serum antibody in birds and amphibians, whereas with regard to structure (the number of exons and inferred positions of disulfide bridges) IgY is more similar to mammalian IgE. Comparison studies of the trans-membrane exons of the ν in duck and the corresponding regions of mammalian γ and ε genes revealed several shared characteristic features [9]. In similar studies of *Xenopus* IgY no particular similarities with either the duck ν or the mammalian γ and ε trans membrane exons were observed. However, conserved residues in the trans-membrane and cytoplasmic domains that are absent in other isotypes were identified [6]. The authors conclude that *Xenopus* IgY, duck IgY and mammalian IgG and IgE probably all share a common precursor and that this IgY-like ancestor arose by a gene duplication of IgM that occurred before the anurans or even the amphibians appeared [6].

Although IgG is the predominant serum Ab in mammals, the role of regulating mast cell degranulation falls mainly on IgE, an antibody that is present at low concentrations. The separation of the effector functions of an IgY-like ancestor into two separate isotypes, IgG and IgE, may therefore have been of selective advantage. Furthermore, the conservation of the overall structure of IgE is strongly indicative of an evolutionary advantage of maintaining the function of the IgE mast cell system. In some mammalian species where mast cells are triggered by both IgE and IgG, as in the guinea pig, a second immunization with an antigen can result in massive anaphylactic reactions [38]. In contrast, in species where mast cells are triggered exclusively by IgE, mast cell degranulation is normally kept under more stringent control. So far, only one IgE isotype has been found in all mammalian species studied, now including the platypus. Hence, restricting regulation of mast cell degranulation to one isotype and to have only one gene for this isotype may have a selective advantage. The potentially very high turnover of IgE raises questions regarding the role of this isotype in the inflammatory response. Does IgE function as a "doorkeeper", scanning the antigen repertoire of the environment and preparing the individual for potentially harmful pathogens [39]? A rapid turnover may then more adequately prepare the individual to deal with its present environment.

Like IgG in other mammals, both of the IgG heavy chains found in platypus (γ1 and γ2) have only three constant domains (Fig. 1 and 3). Hence, the second domain in the Ig ancestral to platypus IgG1 and IgG2 was apparently deleted prior to the evolutionary separation of monotremes from the other mammalian lineages. Alternatively, two or three independent deletions occurred, one in each of the three major mammalian lineages (or one in the common ancestors of marsupials and placental mammals and one in monotremes). However, the high overall structural similarity of the platypus IgG to marsupial and placental IgG strongly favors the first alternative.

In mouse, rat, pig and human the different IgG isotypes are relatively closely related within a single species. In contrast, the sequence divergence between the two platypus IgG isotypes is remarkably high (Fig. 3). In fact, the sequence divergence between the two IgG isotypes in platypus is actually greater than that between the mouse and human IgG isotypes (Fig. 4), indicating that the two platypus IgG isotypes did appear as separate isotypes before the major radiation of placental mammals 60–120 million years ago. Although quite speculative, the distances in the phylogenetic tree presented in Fig. 4 suggest that the platypus IgG isotypes were formed by a gene duplication shortly after the separation of monotremes from the common mammalian ancestor, possibly as long as 150–170 million years ago. The different IgG isotypes in rodents, pigs and humans may instead have arisen from relatively recent duplications within each separate subgroup of eutherians (Fig. 4).

The high degree of sequence divergence between the two platypus IgG isotypes is also reflected in several structural differences. For instance, the position of the cysteine residue involved in the putative disulfide bridge connecting the heavy chain with the light chain differs between the two isotypes. In IgG1, this cysteine residue is located in the N-terminal region of the CH1 domain, whereas in IgG2 it is located in the C-terminal region of the same domain (Fig. 1B and C). To our knowledge, the latter position occurs only in the platypus IgG2 and is thus unique among mammalian IgG molecules. The N-terminal cysteine residue in the CH1 domain of mouse IgG1 has also been lost, but it is not replaced by a cysteine residue in the C-terminal region (Fig. 3). Hence, the heavy chain of mouse IgG1 seems to rely entirely on non-covalent interactions with the light chain, a feature that has been observed in other Ig isotypes as well. For example, disulfide bridges between the heavy and light

chains are absent in the IgM-like Ig and the two low molecular weight Ig isotypes of the bullfrog (*Rana catesbeiana*) [40].

The two platypus IgG isotypes also differ with respect to the number of potential N-linked glycosylation sites. Platypus IgG2 has four potential N-linked glycosylation sites, whereas IgG1 has only one site located in the C2 domain (Fig. 3). However, the lone N-linked glycosylation site in IgG1 is conserved between the two platypus isotypes and among other Ig isotypes as well, including opossum IgG (Fig 3). In fact, this site corresponds to the site in the ε-chain that is conserved among all IgE molecules isolated so far (position 276 in Fig. 2).

In summary, the primary structures of the cloned platypus γ1-, γ2- and ε-chains suggest that all the major evolutionary changes that gave rise to the Ig isotypes expressed by present day mammals emerged prior to the separation of monotremes from marsupials and placental mammals. Furthermore, the presence of three constant domains in both of the highly divergent IgG isotypes found in platypus supports the conclusion that deletion of the second domain of IgG occurred very early in mammalian evolution. The results from the phylogenetic analysis of platypus IgE and IgG are not conclusive with respect to the evolutionary relationship between monotremes and marsupials, indicating that further comparative studies of a large panel of nuclear genes are required to fully resolve the relationship (Fig. 4). However, the platypus IgG sequences and several other unrelated nuclear gene sequences, echidna IgM (Belov et al., submitted for publication), echidna IgA (Belov et al., submitted for publication), protamine, neutropin, α-lactalbumin, β-globin and mannose-6-phosphate/insulin-like growth factor II receptor (M6P/IGF2R) all support the Theria hypothesis [13, 21–24], whereas it is opposed only by platypus IgE and two dental genes, amelogenin and dentin matrix protein 1 [41, 42]. Estimates based on protamine and IgM (Belov et al., submitted for publication) sequences indicate that the monotremes separated from the ancestors of marsupials and placentals about 150–170 million years ago [13], whereas the separation of early mammalian ancestors from early reptiles may have occurred as far back as 310–330 million years ago [14]. Provided that the Theria hypothesis is correct, this leaves us with an evolutionary window for the appearance of modern post-switch isotypes of about 150 million years. Unfortunately, no more living groups of mammals exist today and we cannot obtain information regarding individual molecules from the fossil record. However, further clues may be gained from other molecules that have co-evolved with the various antibody isotypes. For this purpose, we have initiated a study of the high affinity receptor for IgE, FcεRI, in mammals distantly related to placentals. At present, however, we are only able to speculate based on available Ig sequences. Hence, the modern post-switch isotypes may have appeared in the early mammalian ancestors essentially in their present form as early as 310–330 million years ago, shortly after the split from the early reptile lineages.

4 Materials and methods

4.1 Primer design and PCR amplification

Partial cDNA clones of the platypus ε- and γ-chains were isolated by PCR using degenerate primers. The design and sequence of the primers have been described previously [27]. A 10-μl aliquot of the same double-stranded cDNA that was used to generate the platypus spleen cDNA library was used as template in the PCR amplifications. The PCR amplifications were carried out in 100 μl under the following conditions: 50 mM KCl, 10 mM Tris-HCl pH 9.0, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 0.25 mM each of the four nucleotide triphosphates, approximately 1 μg of each of the two degenerate oligonucleotide primers and 1 unit of Taq DNA polymerase (Promega, Madison, WI). The reactions were initiated with a 2-min incubation at 90°C followed by 45 cycles of 40 s at 94°C, 60 s at 52°C and 90 s at 72°C in an MJ Research minicycler (MJ Research, Inc., MC). The PCR products were purified from 1% low-melting agarose gels and cloned into the EcoRI and BamHI sites of Bluescript SK(+/–) (Stratagene, La Jolla, CA). Purified ε- and γ-chain fragments were labeled with ^{32}P by random priming (Megaprime, Amersham Int., GB) and used as probes to screen the platypus cDNA library.

4.2 Isolation of full-length cDNA for platypus IgE, IgG1 and IgG2 (ε-and γ chains)

Total cellular RNA from platypus splenocytes was isolated by the guanidium thiocyanate method as described previously [43] and poly(A)$^+$ RNA was purified using oligo dT-coupled magnetic beads according to the manufacturer's protocol (PolyATtract, mRNA Isolation System II, Promega). Double-stranded cDNA was synthesized using the timesaver cDNA synthesis kit (Pharmacia-Biotech, Uppsala, Sweden). The cDNA was ligated into the single EcoRI site of the λ-gt 10 vector and the recombinant vectors were packaged into phage particles using an *in vitro* packaging system (Stratagene). Approximately 130,000 plaques from this unamplified cDNA library were spread as a monolayer of the *E. coli* C600 Hfl strain with a titer of approximately 17,000 plaques/plate on eight plates. The plaques were transferred to Hybond N+ filters (Amersham Int.). Five of the filters were screened with the ^{32}P-labeled, platypus ε-chain fragment and three filters were screened with the partial γ-chain clone. The filters were washed at high stringency (0.1×SSC,

0.1% SDS). Autoradiography was performed for 24–48 h on Kodak Exomat AR film (Eastman-Kodak Company, Rochester, NY). Positive plaques were purified and the inserts were excised by EcoRI and/or NotI cleavage and subcloned into Bluescript SK(+/–) (Stratagene) or a pGEM-2 vector with a modified multicloning site. The obtained cDNA clones were sequenced from both ends using T7 and T3 or Sp6 primers. Based on the obtained sequences, new sequence primers were designed and used to complete the entire insert sequences. In order to identify additional isotypes, the library was screened with a full-length γ-chain clone (IgG1: 11). Novel signals (negative for the ε- and γ-chain fragment probes) were subcloned and sequenced as above. All nucleotide sequences presented in this study were established by sequencing both strands of the DNA insert.

4.3 Amino acid sequence alignment

The amino acid sequences of the constant regions from a panel of IgE, IgG and IgY sequences were compared using the CLUSTAL W program. This program performs progressive pair-wise comparisons in order of divergence, initiating with the most closely related sequence [44]. The neighbor-joining method was used to construct the evolutionary tree and to estimate the individual branch lengths. A bootstrap tree based on 1,000 independent calculations was constructed using this program (Fig. 4).

The GenBank accession numbers for the various Ig amino acid sequences are listed below.

IgY: duck (X65219) and chicken (X07174).

IgG: platypus 1 (AY055781), 2 (AY055782); rat 2a (M28669), 2b (M28671), 2c (X07189); mouse 1 (J00453), 2a (J00470), 2b (J00461), 3 (J00451); hamster (U17166); rabbit (L29172); sheep 2 (X70983); pig 1 (U03778), 2a (U03779), 2b (U03780), 3 (U03781), 4 (U03782); human 1 (J00228), 2 (J00230), 3 (X03604), 4 (K01316); possum (AF191648) and opossum (AF035195).

IgE: platypus (AF055780), possum (AF157617), opossum (AF035194), rat (K02901), mouse (X01857), sheep (M84356), pig (U96100), horse (U15150), dog (L36872), orangutan (M15399), human (L00022) and chimpanzee (M15398).

4.4 Northern blot analysis

Poly(A)$^+$ RNA derived from platypus splenocytes was isolated as described above. Aliquots of approximately 2 μg of this RNA were loaded in separate lanes on an 1.0% agarose gel containing 0.2 M formaldehyde. The fractionated RNA was blotted onto Hybond-N+ filters using 0.05 M NaOH as transfer solution. The RNA was fixed to the filter by UV cross-linking at 150 mjoule and hybridized overnight at 65°C in 6×SSC, 0.5% SDS, 3×Denhardt's solution, 2 mM EDTA and 100 μg/ml yeast RNA. Random primed ^{32}P fragments of 450–750 bp derived from IgG1, IgG2 and IgE clones, respectively, were used as probes. The filters were washed at high stringency (0.1×SSC, 0.1% SDS) prior to autoradiography. As an internal control of the amount of RNA loaded in each lane, one of the filters was rehybridized with a platypus actin fragment (data not shown).

Acknowledgements: We would like to thank Dr. David Eaker for critically reading the manuscript. This investigation was supported by grants from the Swedish Natural Sciences Research Council (B-BU 9400–301).

References

1 **Du Pasquier, L.,** Origin and evolution of the vertebrate immune system. *APMIS* 1992. **100:** 383–392.

2 **Marchalonis, J. J. and Schluter, S. F.,** A stochastic model for the rapid emergence of specific vertebrate immunity incorporating horizontal transfer of systems enabling duplication and combinational diversification. *J. Theor. Biol.* 1998. **193:** 429–444.

3 **Schluter, S. F., Bernstein, R. M., Bernstein, H. and Marchalonis, J. J.,** 'Big Bang' emergence of the combinatorial immune system. *Dev. Comp. Immunol.* 1999. **23:** 107–111.

4 **Bengtén, E., Wilson, M., Miller, N., Clem, L. W., Pilstrîm, L. and Warr, G. W.,** Immunoglobulin isotypes: structure, function, and genetics. In **Du Pasquier, L. and Litman, G. W.** (Eds.) *Origin and evolution of the vertebrate immune system.* Springer-Verlag, Berlin Heidelberg New York 2000, pp 189–219.

5 **Warr, G. W., Magor, K. E. and Higgins, D. A.,** IgY: clues to the origins of modern antibodies. *Immunol. Today* 1995. **16:** 392–398.

6 **Mussmann, R., Wilson, M., Marcuz, A., Courtet, M. and Du Pasquier, L.,** Membrane exon sequences of the three Xenopus Ig classes explain the evolutionary origin of mammalian isotypes. *Eur. J. Immunol.* 1996. **26:** 409–414.

7 **Marchalonis, J. J.,** Isolation and characterization of immunoglobulin-like proteins of the Australian lungfish (*Neoceratodus forsteri*). *Aust. J. Exp. Biol. Med. Sci.* 1969. **47:** 405–419.

8 **Litman, G. W., Wang, A. C., Fudenberg, H. H. and Good, R. A.,** N-terminal amino-acid sequence of African lungfish immunoglobulin light chains. *Proc. Natl. Acad. Sci. USA* 1971. **68:** 2321–2324.

9 **Magor, K. E., Higgins, D. A., Middleton, D. L. and Warr, G. W.,** One gene encodes the heavy chains for three different forms of IgY in the duck. *J. Immunol.* 1994. **153:** 5549–5555.

10 **Curole, J. P. and Kocher, T. D.,** Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol. Evol.* 1999. **14:** 394–398.

11 **Mussmann, R., Du Pasquier, L. and Hsu, E.,** Is Xenopus IgX an analog of IgA? *Eur. J. Immunol.* 1996. **26:** 2823–2830.

12 **Connolly, J. H., Canfield, P. J., McClure, S. J. and Whittington, R. J.,** Histological and immunohistological investigation of lymphoid tissue in the platypus (*Ornithorhynchus anatinus*). *J. Anat.* 1999. **195:** 161–171.

13 **Retief, J. D., Winkfein, R. J. and Dixon, G. H.,** Evolution of the monotremes. The sequences of the protamine P1 genes of platypus and echidna. *Eur. J. Biochem.* 1993. **218:** 457–461.

14 **Kumar, S. and Hedges, S. B.,** A molecular timescale for vertebrate evolution. *Nature* 1998. **392:** 917–920.

15 **Janke, A., Gemmell, N. J., Feldmaier-Fuchs, G., von Haeseler, A. and Paabo, S.,** The mitochondrial genome of a monotreme-the platypus (*Ornithorhynchus anatinus*). *J. Mol. Evol.* 1996. **42:** 153–159.

16 **Janke, A., Xu, X. and Arnason, U.,** The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria. *Proc. Natl. Acad. Sci. USA* 1997. **94:** 1276–1281.

17 **Penny, D. and Hasegawa, M.,** Molecular systematics. The platypus put in its place. *Nature* 1997. **387:** 549–550.

18 **Kirsch, J. A. and Mayer, G. C.,** The platypus is not a rodent: DNA hybridization, amniote phylogeny and the palimpsest theory. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 1998. **353:** 1221–1237.

19 **Hu, Y., Wang, Y., Luo, Z. and Li, C.,** A new symmetrodont mammal from China and its implications for mammalian evolution. *Nature* 1997. **390:** 137–142.

20 **Luo, Z. X., Cifelli, R. L. and Kielan-Jaworowska, Z.,** Dual origin of tribosphenic mammals. *Nature* 2001. **409:** 53–57.

21 **Kullander, K., Carlson, B. and Hallbook, F.,** Molecular phylogeny and evolution of the neurotrophins from monotremes and marsupials. *J. Mol. Evol.* 1997. **45:** 311–321.

22 **Messer, M., Weiss, A. S., Shaw, D. C. and Westerman, M.,** Evolution of the monotremes: phylogenetic relationship to marsupials and eutherians, and estimation of divergence dates based on α-lactalbumin amino acid sequences. *J. Mamm. Evol.* 1998. **5:** 95–105.

23 **Lee, M. H., Shroff, R., Cooper, S. J. and Hope, R.,** Evolution and molecular characterization of a beta-globin gene from the Australian Echidna *Tachyglossus aculeatus* (*Monotremata*). *Mol. Phylogenet. Evol.* 1999. **12:** 205–214.

24 **Killian, J. K., Buckley, T. R., Stewart, N., Munday, B. L. and Jirtle, R. L.,** Marsupials and eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution. *Mamm. Genome* 2001. **12:** 513–517.

25 **Springer, M. S., DeBry, R. W., Douady, C., Amrine, H. M., Madsen, O., de Jong, W. W. and Stanhope, M. J.,** Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol. Biol. Evol.* 2001. **18:** 132–143.

26 **Archer, M., Flannery, T. F., Ritchie, A. and Molnar, R. E.,** First Mesozoic mammal from Australia – an early Cretaceous monotreme. *Nature* 1985. **318:** 363–366.

27 **Aveskogh, M. and Hellman, L.,** Evidence for an early appearance of modern post-switch isotypes in mammalian evolution; cloning of IgE, IgG and IgA from the marsupial *Monodelphis domestica*. *Eur. J. Immunol.* 1998. **28:** 2738–2750.

28 **Aveskogh, M., Pilstrom, L. and Hellman, L.,** Cloning and structural analysis of IgM (μ chain) and the heavy chain V region repertoire in the marsupial *Monodelphis domestica*. *Dev. Comp. Immunol.* 1999. **23:** 597–606.

29 **Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S., and Foeller, C.,** Sequences of proteins of immunological interest, Fifth Edn. U.S. Department of Health and Human Services, Bethesda, MD 1991.

30 **Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. and Winter, G.,** Structural repertoire of the human VH segments. *J. Mol. Biol.* 1992. **227:** 799–817.

31 **Vu, K. B., Ghahroudi, M. A., Wyns, L. and Muyldermans, S.,** Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol. Immunol.* 1997. **34:** 1121–1131.

32 **Roux, K. H., Greenberg, A. S., Greene, L., Strelets, L., Avila, D., McKinney, E. C. and Flajnik, M. F.,** Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. *Proc. Natl. Acad. Sci. USA* 1998. **95:** 11804–11809.

33 **Saini, S. S., Allore, B., Jacobs, R. M. and Kaushik, A.,** Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *Eur. J. Immunol.* 1999. **29:** 2420–2426.

34 **Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A. and Hamers, R.,** Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng.* 1994. **7:** 1129–1135.

35 **Johansson, J., Aveskogh, M., Munday, B. and Hellman, L.,** Heavy chain V region diversity in the duck-billed platypus (*Ornithorhynchus anatinus*): long and highly variable complementarity-determining region 3 compensates for limited germline diversity. *J. Immunol.* 2002, **168:** 5155–5162.

36 **Belov, K., Harrison, G. A., Miller, R. D. and Cooper, D. W.,** Molecular cloning of the brushtail possum (*Trichosurus vulpecula*) immunglobulin E heavy chain constant region. *Mol. Immunol.* 1999. **36:** 1255–1261.

37 **Hasemann, C. A. and Capra, J. D.,** Immunoglobulins: structure and function. In **Paul, W. E.** (Ed.) *Fundamental Immunology.* Raven Press, New York 1989, pp 226.

38 **Andersson, P.,** Antigen-induced bronchial anaphylaxis in actively sensitized guinea-pigs. Pattern of response in relation to immunization regimen. *Allergy* 1980. **35:** 65–71.

39 **Hellman, L.,** Genes for immunoglobulin E: structure, rearrangement and expression. Doctoral thesis. ISSN 0345-0058, ISBN 91-554-1721-3. The Department of Microbiology, Uppsala University, Uppsala 1985.

40 **Mikoryak, C. A. and Steiner, L. A.,** Noncovalent association of heavy and light chains in *Rana catesbeiana* immunoglobulins. *J. Immunol.* 1984. **133:** 376–383.

41 **Toyosawa, S., O'Huigin, C., Figueroa, F., Tichy, H. and Klein, J.,** Identification and characterization of amelogenin genes in monotremes, reptiles, and amphibians. *Proc. Natl. Acad. Sci. USA* 1998. **95:** 13056–13061.

42 **Toyosawa, S., O'Huigin, C. and Klein, J.,** The dentin matrix protein 1 gene of prototherian and metatherian mammals. *J. Mol. Evol.* 1999. **48:** 160–167.

43 **Chomczynski, P. and Sacchi, N.,** Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 1987. **162:** 156–159.

44 **Thompson, J. D., Higgins, D. G. and Gibson, T. J.,** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994. **22:** 4673–4680.

Correspondence: Lars Hellman, Department of Cell and Molecular Biology, University of Uppsala, The Biomedical Center, Box 596, S-751 24 Uppsala, Sweden
Fax: +46–18–471–4434
e-mail: Lars.Hellman@icm.uu.se